TESIS DE LA UNIVERSIDAD
DE ZARAGOZA

2024

166

Jorge Navarro López

Dinámica social de las emociones: Redes sociales y patrones emocionales colectivos analizados mediante técnicas de machine learning e Inteligencia Artificial

Director/es
Aguarón Joven, Juan Alfredo
González Álvarez, María Asunción

ISSN 2254-7606

Prensas de la Universidad
Universidad Zaragoza

Universidad de Zaragoza
Servicio de Publicaciones

ISSN 2254-7606

Universidad Zaragoza
1542

Tesis Doctoral

DINÁMICA SOCIAL DE LAS EMOCIONES: REDES SOCIALES Y PATRONES EMOCIONALES COLECTIVOS ANALIZADOS MEDIANTE TÉCNICAS DE MACHINE LEARNING E INTELIGENCIA ARTIFICIAL

Autor

Jorge Navarro López

Director/es

Aguarón Joven, Juan Alfredo
González Álvarez, María Asunción

UNIVERSIDAD DE ZARAGOZA
Escuela de Doctorado

2023

Universidad Zaragoza
1542

Tesis Doctoral

Dinámica social de las emociones: Redes sociales y patrones emocionales colectivos analizados mediante técnicas de machine learning e Inteligencia Artificial

Autor

Jorge Navarro López

Director/es

María Asunción González Álvarez
Juan Alfredo Aguarón Joven

Facultad de Economía y Empresa

2023

UNIVERSIDAD DE ZARAGOZA

FACULTAD DE ECONOMÍA Y EMPRESA

DEPARTAMENTO DE ECONOMÍA APLICADA

Dinámica social de las emociones:

Redes sociales y patrones emocionales colectivos analizados mediante técnicas de *machine learning* e Inteligencia Artificial

Autor:

Jorge Navarro López

Directores:

María Asunción González Álvarez

Juan Alfredo Aguarón Joven

Doctorado en Economía

Facultad de Economía y Empresa

Universidad de Zaragoza

2023

A mi mujer, mis hijas, mi madre y mi hermano, apoyos incondicionales

Agradecimientos:

- *A la Dra. María Asunción González Álvarez, por su apoyo constante y supervisión.*
- *Al Dr. Juan Alfredo Aguarón Joven, por su apoyo constante y supervisión*
- *Al Dr. José María Moreno Jiménez, por su apoyo y buena acogida dentro del Grupo de Investigación GDMZ.*
- *Al Dr. Pedro C. Marijuán, por estimular mi interés por la investigación y por su continua colaboración y apoyo.*
- *A mis compañeros y compañeras del GDMZ, por su apoyo y porque trabajar así es un placer.*
- *Al Gobierno de Aragóny los fondos FEDER por su soporte económico mediante el grupo de investigación 'Grupo Decisión Multicriterio Zaragoza' (S35-20R) y el proyecto 'Participación Ciudadana Cognitiva y Decisiones Públicas. Aplicaciones Sociosanitarias' (Ref. LMP35_21).*
- *Al Ministerio de Ciencia e Innovación de España por su soporte económico mediante el proyecto "Codecisión Cognitiva y Colaborativa. Aplicaciones en Salud" (Ref.: PID2022-139863OB-I00).*
- *A Laura, mi amiga, compañera y esposa, que tanta paciencia has tenido con "mi tesis". Gracias por tu apoyo incondicional!!!*
- *A mis hijas Sofía y Blanca, las princesas de mi vida.*
- *Y por último, a mi madre y mi hermano, que tanto apoyo me han mostrado siempre con éste y todos mis proyectos.*

La presente Tesis doctoral, que se ha realizado bajo la modalidad de compendio de publicaciones, incluye los siguientes trabajos:

- Navarro, J., Cañete, M., Olivera, F. J., Gil-Lacruz, M., Gil-Lacruz, A., & Marijuán, P. C. (2022). The Cost of Loneliness: Assessing the Social Relationships of the Elderly via an Abbreviated Sociotype Questionnaire for inside and outside the Clinic. *International Journal of Environmental Research and Public Health*, 19(3), 1253.

  https://doi.org/10.3390/ijerph19031253

  Factor de impacto SJR: 0,828 (Q2)

- Navarro, J.* & Marijuán, P. C. (2023). Natural intelligence and the 'economy' of social emotions: A connection with AI sentiment analysis. *Biosystems*, 233, 105039.

  https://doi.org/10.1016/j.biosystems.2023.105039

  Factor de impacto JCR: 1.6 (Q3)

- Turón, A., Altuzarra, A., Moreno-Jiménez, J. M., & Navarro, J.* (2023). Evolution of social mood in Spain throughout the COVID-19 vaccination process: a machine learning approach to tweets analysis. *Public health*, 215, 83-90.

  https://doi.org/10.1016/j.puhe.2022.12.003

  Factor de impacto JCR: 5,2 (Q1)

- Navarro, J., Pina, J. U., Mas, F. M., & Lahoz-Beltra, R. (2023). Press media impact of the Cumbre Vieja volcano activity in the island of La Palma (Canary Islands): A machine learning and sentiment analysis of the news published during the volcanic eruption of 2021. *International Journal of Disaster Risk Reduction*, 91, 103694.

  https://doi.org/10.1016/j.ijdrr.2023.103694

  Factor de impacto JCR: 5,0 (Q1)

RESUMEN

La presente Tesis, realizada mediante un compendio de publicaciones, representa la línea de investigación seguida por el Autor a lo largo de dos etapas distintas. En ambas etapas, sin embargo, la temática de investigación ha permanecido coherente: las redes sociales, los lazos interindividuales, y las emociones que acompañan a las dinámicas de socialización. Lo que ha variado es la metodología de análisis. En primer lugar, a través del trabajo de campo y la creación de un nuevo test sobre la desaparición de las redes sociales del individuo (el grave problema de detectar la soledad no deseada entre los mayores) se ha procedido a analizar la conexión entre los estratos de relación social (lazos fuertes versus lazos débiles) junto con la respectiva presencia de emociones bien diferenciadas (primarias versus secundarias) en el contexto de los procesos adaptativos del "sociotipo". Paralelamente, todo ello se ha continuado a otro nivel mediante el uso de herramientas de inteligencia artificial para el análisis de sentimientos (*sentiment analysis*), combinando estas técnicas de manera novedosa con *machine learning* y análisis estadístico multivariante. Estas técnicas se han aplicado, entre otros trabajos, al análisis de intercambios masivos de mensajes en las redes sociales durante la reciente pandemia y al estudio de compilaciones de noticias publicadas sobre una catástrofe natural, como la reciente erupción del volcán Cumbre Vieja en las Islas Canarias. Los resultados obtenidos a lo largo de esta línea de investigación pueden contribuir, por un lado, a mejorar la detección de la soledad en personas mayores y a clarificar los procesos emocionales en las relaciones sociales en general. Por otro lado, y muy especialmente, al seguimiento de las consecuencias en la opinión pública de las decisiones y políticas adoptadas tanto en situaciones de graves desafíos sanitarios como frente a catástrofes naturales. Indudablemente, la repercusión multidisciplinar de este tipo de estudios de análisis de sentimientos, que complementan la teoría de decisión y el análisis de riesgos, es cada vez mayor y abarca múltiples ramas de las ciencias sociales, la economía, las ciencias políticas y las ciencias de la comunicación.

ÍNDICE

1.INTRODUCCIÓN

1.INTRODUCCIÓN

1.1. Línea conceptual seguida

La línea de investigación que sirve de base a este proyecto de doctorado se titula: **"Dinámica social de las emociones: redes sociales y patrones emocionales colectivos analizados mediante técnicas de *machine learning*e Inteligencia Artificial."**

Esta línea continúa, por un lado, los trabajos de investigación realizados por el doctorando durante su etapa en el grupo de Bioinformación del Instituto Aragonés de Ciencias de la Salud (IACS), que estaban centrados en el análisis de los sistemas basados en el procesamiento de la información, desde los sistemas biológicos a los sistemas nerviosos y los flujos de la información y la comunicación en nuestras propias sociedades. En la dinámica de las interacciones sociales, como se discute en el presente trabajo, las emociones y sus procesos colectivos aparecen jugando un papel central, no siempre bien entendido ni reconocido en las disciplinas tradicionales. Con la presente Tesis lo que pretendemos es clarificar y avanzar en ese entendimiento, necesariamente multidisciplinar, así como destacar algunas importantes consecuencias (sociales, económicas y culturales) derivadas de dichas dinámicas emocionales. Por ejemplo, los problemas derivados del creciente aislamiento y soledad de los mayores, que abordamos en uno de los trabajos presentados. Y por otro lado, mediante la aplicación de técnicas de *machine learning* e Inteligencia Artificial (*sentiment analysis*), pretendemos analizar con una mayor extensión y riqueza de detalle los contenidos emocionales que acompañan a la comunicación masiva a través de las nuevas tecnologías en las redes sociales. Es importante que con estas técnicas hayamos podido analizar los cambios en la opinión pública, en particular durante la pandemia COVID-19, como respuesta a las decisiones políticas adoptadas por los diferentes niveles de la Administración. O que hayamos hecho un análisis de los contenidos emocionales y los cambios de opinión en los medios de comunicación durante la reciente erupción del volcán Cumbre Vieja en las Islas Canarias

Para entender la dinámica social de las emociones es útil acudir primero al concepto del "sociotipo", ampliamente desarrollado en diversos artículos y varios tests a los que nos referiremos luego (Marijuán et al., 2017; Marijuán and Navarro, 2020; Navarro et al., 2022). *El sociotipo representa nuestra socialidad adaptativa,* el conjunto relacional compuesto por los círculos vinculares de la familia y los parientes, los amigos y los conocidos. Necesitamos reacciones emocionales instintivas para lograr nuestra adecuación individual en esos grupos y para lograr construir y mantener nuestro propio sociotipo, y también para lograr el desarrollo de la inteligencia social colectiva en el grupo. Al respecto, una primera pregunta que este autor ha intentado responder es, ¿qué ocurre cuando no hay grupo? En otras palabras, ¿cómo nos ayuda la idea del sociotipo a detectar y cualificar la **soledad no deseada**? Singularmente uno de los tests del sociotipo (el "sociotipo abreviado", de sólo cuatro preguntas), desarrollado por este autor en cooperación con el Grupo de Psicogeriatría del Hospital San Jorge de Huesca, ha sido dedicado a establecer el coste social de las "no relaciones", es decir a poder detectar y evaluar con la mayor simplicidad posible el grave problema contemporáneo de la soledad no deseada, derivado de la ausencia de redes sociales efectivas alrededor del individuo (Navarro et al., 2022). El impacto directo de la soledad en la salud mental y física, así como la carga asistencial y social que provoca, ya ha sido destacada por varios investigadores en campos tan diversos como la economía social (Granovetter, 2018; Larg and Moss, 2011; McPherson et al., 2006; Mihalopoulos et al., 2020; Putnam, 2000), la psicología social (Qualter et al., 2015; Shankar et al., 2011; Shor et al., 2013), la salud mental y la psiquiatría (Booth, 2000; Cacioppo and Cacioppo, 2014; Grant et al., 2009). Hay que resaltar que, desde la pandemia COVID-19, la soledad no deseada se ha convertido en una preocupación social, económica y política aún más importante (WHO, 2020). Por ello, la necesidad de continuar con el desarrollo de **herramientas sencillas** para su detección, especialmente por los facultativos de atención primaria y en las consultas geriátricas, es realmente apremiante (Navarro et al., 2022).

Pero a la vez necesitamos **herramientas complejas** para abordar las dinámicas emocionales con la mayor extensión y profundidad posibles. De ahí que en su nueva etapa investigadora este autor haya profundizado en el tema de las emociones colectivas, analizándolas ahora mediante técnicas de inteligencia Artificial. Es un tema que se suele plantear tradicionalmente como una dinámica propiamente psicológica a base de unas cuantas emociones básicas que resultan más fáciles de expresar tanto a través del lenguaje verbal como de las expresiones faciales y el lenguaje corporal. Aquí lo abordamos de una manera distinta que creemos más extensa y más profunda, tanto teórica como analíticamente. Y lo hacemos mediante las mencionadas técnicas de *machine learning* y *sentiment analysis*, para evidenciar así a partir de datos reales la emergencia de patrones emocionales colectivos en los mensajes intercambiados en las redes sociales.

1.2. Las nuevas técnicas de Inteligencia Artificial

El actual proyecto de investigación, en sus aspectos más empíricos, está asimismo alineado con la actividad investigadora del Grupo de Investigación GDMZ de la Facultad de Economía y Empresa (al que pertenece el autor desde que comenzó la Tesis). Se trata de dotar de rigor científico la toma de decisiones en situaciones altamente complejas caracterizadas por la existencia de múltiples actores, escenarios y criterios (tangibles e intangibles). Por ello esta investigación se centra en el desarrollo de herramientas decisionales (analíticas e informáticas) que mejoren la eficiencia (hacer correctamente las cosas) y la eficacia (alcanzarlas metas fijadas) de las diferentes etapas del proceso decisional y, fundamentalmente, la efectividad (hacer lo correcto) del sistema.

Cuando observamos la emergencia de patrones emocionales colectivos, también puede constatarse un cierto grado de inteligencia social en relación con los comportamientos de los individuos agregados en entidades e instituciones supraindividuales para conseguir, a través de la influencia de sus redes de comunicación, tanto hacer avanzar sus propios intereses vitales como los de la comunidad. Es decir, la toma de decisiones colectivas que van acompañadas o precedidas, o seguidas *a posteriori*, por toda una nube de intercambios en los que podemos detectar analíticamente la presencia y la dominancia de determinadas emociones netamente sociales.

El argumento central es que el papel de las emociones en el ámbito social es el de preceder y dotar de "sentido común" a las diversas situaciones del individuo en el seno del colectivo social, orientando eficazmente sus decisiones y acciones desde un punto de vista adaptativo. En particular, la pandemia COVID-19 ha proporcionado un escenario fundamental donde la opinión pública ha ido respondiendo de forma casi inmediata a la toma de decisiones sobre vacunación y salud pública por parte de las distintas autoridades. Lo hemos estudiado en una de las publicaciones de esta Tesis (Turón et al., 2023), y no sólo hemos estudiado la emergencia de los patrones emocionales colectivos mediante técnicas de *sentiment analysis*, sino que también hemos podido correlacionar los estados de ánimo colectivos con las distintas novedades y medidas públicas adoptadas en relación con la pandemia. Otro trabajo aquí presentado, hasta cierto punto similar, en cooperación con la Universidad Complutense, tiene que ver también con el empleo de *sentiment analysis* para evaluar el impacto en los medios de prensa nacionales e internacionales de sucesos de alta relevancia social, en concreto la erupción del volcán Cumbre Vieja (Navarro et al., 2023).

Por diversos motivos que luego discutiremos con detalle, la inteligencia artificial que durante largas décadas ha sido exclusivamente “racional” se ha ido ocupando de las emociones con cada vez más interés. Las reacciones emocionales, como las más recientes investigaciones que luego veremos atestiguan, han resultado cruciales, tanto en la evolución de nuestra especie como en la emergencia de su cerebro social y su propia inteligencia social. Incluso actualmente se discute en la propia inteligencia artificial que sin las emociones no hay valoraciones individuales ni una auténtica racionalidad en el comportamiento individual. Como hemos planteado en nuestros trabajos (Navarro and Marijuán, 2023, 2022), las emociones son consustanciales a la organización del procesamiento individual y social de la información, incluso ya en el sistema biológico complejo, dotado de sistema nervioso central, donde emerge la dinámica de la inteligencia adaptativa, que aquí desarrollamos través de la noción de la **“Inteligencia Natural”**. Se argumenta que no resulta viable entender el sistema emocional humano (o de cualquier otra especie animal) sin atender a su continuo despliegue de procesamientos inteligentes que adaptan el individuo a su medio, a su nicho, que en el caso humano es esencialmente social. Para el estudio *in extenso* de estas emociones estamos utilizando en los trabajos presentados, como ya se ha dicho, herramientas de inteligencia artificial, *sentiment analysis*, integrando *machine learning*. Evidentemente cubren aspectos bastante parciales, y deben entenderse como un *work in progress* que sirve para marcar una dirección prometedora de investigación con impacto directo en ámbitos muy diversos: mediáticos, económicos, políticos, culturales, etc.

Esta mezcla o convergencia de las dos clases de inteligencia, la natural y la artificial, de la que aún estaríamos lejos es fundamental para avanzar en el estudio de la dinámica social de las emociones. Lo vamos a argumentar con detalle en sección que sigue.

1.3. Rediscutir la naturaleza de las emociones

El planteamiento de esta tesis, netamente multidisciplinar, lleva a discusiones dentro de muy distintas disciplinas sobre temas que tradicionalmente han generado una amplia controversia y que están lejos de ser resueltos satisfactoriamente. En lo que sigue vamos a discutir primero sobre la naturaleza de las emociones, y ello es debido a que como veremos el tratamiento de las emociones sociales, y de las emociones en general, resulta excesivamente simplificado en el análisis de sentimientos. Realizaremos primero un examen preliminar de las emociones y los esquemas hoy día más utilizados, luego trataremos sobre las emociones más claramente sociales y su relación con las diversas categorías de lazos sociales discutidos en el sociotipo, y finalmente propondremos una posible línea de investigación en análisis de sentimientos que pondría de manifiesto los principales desencadenantes de las distintas emociones.

Para empezar a discutir sobre las emociones conviene adelantar un esquema general, el conocido diagrama cartesiano de la emoción (Posner et al., 2005; Russell, 1980; Russell and Barrett, 1999). Es el "modelo circumplejo", que cuenta con la *valencia* y la *excitación* como las dos dimensiones fundamentales del espacio emocional. Véase la figura 1. Este modelo, que ha tenido una amplia repercusión, tiene la ventaja de situar muy variadas emociones en lugares significativamente conectados entre sí. De hecho, de este modelo se han derivado numerosos mapas gráficos de emociones, a menudo introduciendo una tercera dimensión que suele ser la subsiguiente actuación de acercamiento/evitación, o bien el tiempo. El modelo también puede proporcionar una comprensión visual de las trayectorias emocionales cuando las coordenadas de valencia y activación de los sujetos se modifican en función de la evolución en el tiempo de los estados mentales. Además, la persistencia del sujeto en algún estado emocional equivaldría a un desplazamiento permanente del origen de las coordenadas, de modo que se necesitaría comparativamente más -o menos- activación, o un incremento mayor -o menor- en cuanto a valencia. La permanencia en el tiempo de algunos estados emocionales se considera a menudo como un caso de "sentimiento", al menos en la forma en que se practica actualmente el análisis del sentimiento en IA, como se verá más adelante.

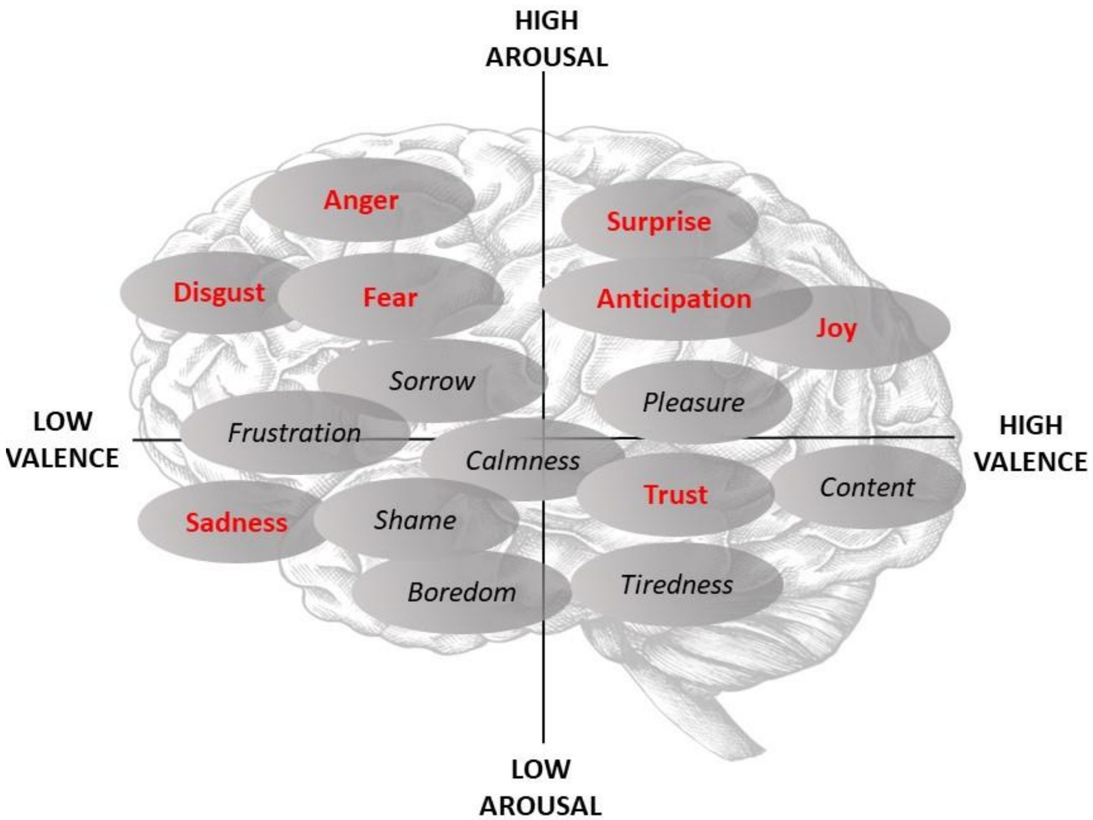

Figura 1. El modelo circumplejo: Esta representación cartesiana (excitación versus valencia) incluye las ocho emociones básicas/primarias mencionadas en el texto principal, en rojo, y otras ocho emociones secundarias, más sociales (en cursiva). Fuente: Elaboración propia.

Otro modelo relevante a considerar se debe a Robert Plutchik (Plutchik, 1980), conocido como la "rueda de las emociones", con ocho emociones primarias agrupadas en influencias positivas frente a negativas y capaces de combinarse para formar díadas y tríadas emocionales. Este modelo también ha sido seguido por muchos profesionales y orientadores sobre trastornos de la personalidad y superación personal. Pero también se han desarrollado muchas otras listas y clasificaciones, y el número de emociones enumeradas no ha dejado de aumentar, así como sus posibles combinaciones (Parrott, 2001). Por ejemplo, seis ejes de emociones cada uno con otros seis rangos graduales dan un total de 36 emociones (Kort et al., 2001), y el Libro de las Emociones Humanas contiene un total de 154 emociones y sentimientos identificables en diferentes culturas (Watt Smith, 2021). Es interesante que las seis grandes emociones de Ekman fueran complementadas posteriormente con otras 16 por sus colaboradores en la investigación

(Ekman, 1999), la mayoría de ellas sociales y no necesariamente expresadas en los músculos faciales.

En realidad, uno de los principales problemas de los teóricos de las emociones, al menos para la investigación vinculada a plataformas comerciales y a la IA más aplicada, no es la recopilación de posibles listas de emociones, sino tratar de establecer un sólido correlato emocional para las situaciones sociales más frecuentes de la vida cotidiana, ya sea frente a una pantalla o en relaciones cara a cara, o en contextos grupales. Lo que predomina, casi abrumadoramente, son los enfoques de fuerza bruta basados en *big data*, que son útiles para fines de marketing directo, pero que en absoluto bastan para desarrollar una perspectiva coherente de las emociones. Como ejemplo revelador, la risa y el llanto, estados emocionales tan básicos en los grupos humanos cercanos y fundamentales respectivamente para apoyar la creación de vínculos o lazos sociales y para mitigar su destrucción (Marijuan and Navarro, 2010; Navarro et al., 2016, 2014), siguen ausentes en casi todas las recopilaciones de emociones. Esta ausencia nos indica que los aspectos relacionales de importancia podrían no estar bien resueltos aún en los enfoques convencionales de la emoción. El enigmático papel de la risa a lo largo de todas las etapas de la vida humana parece consistir en una herramienta indirecta pero altamente eficiente para la creación de vínculos interindividuales a través del refuerzo sináptico (Navarro et al., 2016).

De acuerdo con lo planteado en uno de los artículos presentados en esta Tesis (Navarro and Marijuán, 2023), lo que la inteligencia natural nos sugeriría es la necesidad de mirar de cerca el ciclo vital humano (o mejor, el curso vital), buscando los patrones de detección fijos y flexibles y los patrones superestructurales que son capaces de agitar nuestros recursos de procesamiento y reorientar nuestros mecanismos cognitivos precisamente mediante herramientas emocionales a lo largo del curso vital. Dado que nuestra adaptación fundamental es a un entorno social muy rico y estructurado. Debemos intentar sintetizar cuáles podrían ser los principales patrones estructurales y superestructurales que desencadenan nuestra panoplia de emociones en la adaptación a un entorno social muy complejo y con notables diferencias a lo largo del curso vital—y por supuesto con respecto a la edad, al género, la clase social, la cultura, etc. Y aquí conectamos de lleno con la ya comentado acerca del sociotipo.

1.4. Ejes fundamentales de las emociones sociales

En una primera aproximación a las emociones sociales, podemos considerar que las "seis grandes" emociones tradicionalmente discutidas por los teóricos (Barrett, 2006; Ekman, 1999, 1992) son las más importantes en cuanto a sus expresiones faciales y corporales (tristeza, felicidad, miedo, ira, sorpresa y asco), pero esto no significa que sean las más comunes o significativas en nuestra vida cotidiana dentro de los diferentes entornos sociales o en la comunicación online. Más bien podemos encontrar con mayor frecuencia una serie de sentimientos duales y reacciones emocionales vinculadas a situaciones grupales como: exclusión vs. inclusión, simpatía vs. antipatía, admiración vs. envidia, recompensa vs. castigo, irritabilidad vs. calma, excitación vs. compostura, etc. Consideraremos algunas de estas emociones más sociales dentro del nuevo marco que estamos explorando en *sentiment analysis*. Efectivamente, los trabajos actuales de IA sobre análisis de sentimientos a través de léxicos están intentando analizar algunas de estas reacciones duales, procesándolas regularmente a través de diferentes procedimientos (Turón et al., 2023). La interrelación de los textos de las redes sociales con las emociones es también una cuestión esencial para las plataformas comerciales (Zuboff, 2019).

Tratando de separar los diferentes dominios emocionales pertenecientes a los principales escenarios relacionales de nuestra sociedad, podríamos distinguir las siguientes modalidades descritas en nuestro estudio del sociotipo: (i) supervivencia y automantenimiento; (ii) sexo y núcleo familiar; (iii) amigos, colegas y conocidos en general; y (iv) ultrasocialidad e identidades colectivas.

En cuanto a los impulsos de la supervivencia y el automantenimiento, (i) estarían atendidos por las emociones básicas ya mencionadas, sea cual sea el número que consideremos. (Por ejemplo, las seis básicas de Ekman: tristeza, felicidad, miedo, ira, sorpresa y asco). Y en cuanto al sexo, el matrimonio y la vida familiar, (ii) también estarían cubiertos por estas emociones básicas, pero además por algunas nuevas emociones específicas ligadas a fuertes procesos vinculares, "casi" exclusivamente humanos (por ejemplo: amor, afecto, lujuria, juego, risa, curiosidad). Estas seis más seis podrían considerarse nuestras emociones "primarias" para establecer vínculos fuertes. El núcleo esencial de nuestro sociotipo.

A continuación, podríamos agrupar otro sector, el de los "amigos", que en cierto modo es intermedio entre las relaciones nucleares y los colegas de trabajo y conocidos en general, considerándolos a todos juntos bajo la etiqueta de "interindividual" (iii). En este ámbito interindividual del sociotipo tenemos el instinto de crear vínculos sociales de naturaleza más débil, más numerosos y maleables, lo que implica frecuentes inclusiones/exclusiones. En este dominio de vínculos débiles, debemos mantener nuestra reputación e imagen personal, debemos cooperar para lograr nuestros mejores intereses con conflictos ocasionales con otros individuos, e instintivamente acatamos reglas relacionales estrictas, incluso a edades muy tempranas (Tomasello, 2019). Este es el territorio genuino donde se despliegan las emociones moralistas descritas por Trivers como estrategias conductuales espontáneas en el juego de la reciprocidad (Pinker, 2009; Tomasello, 2019; Trivers, 1985). Serían otras seis emociones "secundarias" que aparecerían, digamos, en paralelo a las emociones primarias que acabamos de mencionar para los vínculos fuertes. Encontraríamos para los vínculos débiles: resentimiento, agrado, gratitud, simpatía, culpa y vergüenza. En este punto deberíamos recordar la importante diferencia en ciencias sociales entre vínculos o enlaces fuertes y los débiles, y la centralidad de estos últimos en el establecimiento de actividades comerciales y económicas (Granovetter, 1973). Los vínculos débiles y sus emociones asociadas serían el principal soporte del civismo. Así, podemos establecer una interesante correspondencia de los vínculos con las emociones: las emociones primarias y secundarias se encargarían respectivamente de crear y mantener los vínculos fuertes, en el caso de las primeras, y los vínculos débiles las segundas.

En este escenario interindividual, proponemos que en una analogía con los planteamientos tridimensionales del modelo circumplex, aparecerían tres ejes distintivos o "patterning" altamente frecuentes para el desencadenamiento de las emociones, precisamente donde surgen la mayoría de los conflictos grupales. Consistirían en: confianza (cooperación) frente a desconfianza (conflicto); superioridad (arrogancia) frente a inferioridad (humillación); e inclusión (aceptación) frente a exclusión (rechazo), lo cual es muy significativo en muchos casos en los que ser marginado en un grupo equivale a tener un futuro realmente sombrío. Y una condición adicional a considerar tiene que ver con la distancia cognitiva, familiar (cercano) vs. no familiar (lejano), que es altamente relevante en lo que se refiere a nuestro procesamiento "automático" de minimización y la subsiguiente respuesta emocional a los patrones coligados presentes en estos ejes anteriores (Collins, 1991; Collins and Marijuán, 1997). A partir de ahí, las respectivas coordenadas de los diferentes patrones en este espacio multidimensional, convenientemente transformadas por la condición 'familiar' vs. 'no familiar', pondrían en acción diferentes emociones, bien de fuerte vinculación -primarias-, bien otras de naturaleza interindividual -secundarias- más relacionadas con nuestra inclinación prosocial ampliada.

Y también está el importantísimo fenómeno de la ultrasocialidad (iv). Históricamente, nos hemos embarcado en identidades colectivas de naturaleza y tamaño muy variables. Nuevos determinantes como lo común (unidad) frente al individualismo (discordia), la libertad (tolerancia) frente a la opresión (intolerancia) y la igualdad (equidad, justicia) frente a la desigualdad (injusticia) representan otras dimensiones importantes o ejes de patrones para asignar nuestras respuestas emocionales en relación con las identidades colectivas. De algún modo, se hace eco del eslogan político clásico de "Liberté, Égalité, Fraternité". Pero evolutivamente, el tiempo de la historia para este fenómeno ultrasocial ha sido demasiado corto. Muy probablemente, la utrasocialidad ha estado utilizando los recursos emocionales ya presentes en el Homo sapiens. Así, habría cooptado una mezcla de emociones básicas y secundarias en el surgimiento y mantenimiento de las nuevas estructuras sociales, sobre todo a través de desarrollos políticos, religiosos y culturales. Sin embargo, como en el caso de la lectura (Dehaene, 2009) bien podría ser que el proceso de socialización de los individuos, su "educación", provoque la emergencia de emociones genuinamente nuevas derivadas de la combinatoria entre reacciones emocionales previas, con reacciones como: euforia, admiración, asombro, adhesión, sincronización, unión; así como rechazo, hostilidad, xenofobia...

En resumen, este enfoque de las emociones sociales basado en la inteligencia natural y en nuestra naturaleza social que se condensa en el sociotipo está proponiendo un tipo de exploración diferente. No creemos, por el momento, que una enumeración detallada de las emociones prosociales tomadas de forma aislada sea factible, ni siquiera interesante. Nuestro enfoque alternativo -complementario- basado en la detección de ejes de patrones asociados a los diferentes dominios relacionales, más que oponerse a los campos de la IA, podría desarrollarse en una fructífera cooperación con ellos. Las ideas básicas sobre el análisis de sentimientos que se exponen a continuación podrían proporcionar algunas pistas sobre cómo contribuir de manera significativa, de forma empírica, a refinar y desarrollar las presentes sugerencias.

1.5. Encaje de las emociones en la IA: análisis de sentimientos

Las emociones nunca han sido ajenas a la IA. A pesar de la primera oleada de enfoques pragmáticos y racionalistas, las emociones y las métricas emocionales se incorporaron a la IA relativamente pronto. Desde los rudimentarios "sentómetros" de Manfred Clynes en los años 80 (Clynes, 1988), representantes de la nueva ola que estaba por llegar, hasta la investigación sobre "computación afectiva" encabezada por Rosalind Picard en los años 90 (Picard, 1997). Más tarde, con el creciente interés empresarial por la "economía de la atención", surgieron diferentes sistemas de detección de emociones y patentes por parte de Facebook, Affectiva, Emoshape y otros, incluyendo el diseño de microchips específicos (Zuboff, 2019). Un referente importante fueron las "seis grandes" emociones básicas de Eksman, que como ya hemos comentado se convirtieron en 16 en la investigación de Facebook. Una de las principales orientaciones fue hacia la detección visual automática de emociones en imágenes, así como la identificación de múltiples combinatorias emocionales. En otra dirección, se abordaron los "sociómetros" de Alex Pentland para cubrir métricas importantes sobre las relaciones sociales: interacciones, procesos de vinculación, jerarquías reales, etc—fácilmente medibles a través de datos de *wearables* y móviles ad hoc (Pentland, 2014). Además, el auge de las redes sociales, con su enorme reguero de imágenes, vídeos y textos, ha promovido nuevos tipos de enfoques de IA. Uno de ellos, el análisis de sentimientos, es el eslabón de investigación que tomamos como vía potencial para conectar la inteligencia natural con la inteligencia artificial en lo que respecta a las emociones sociales (a propósito, dejamos de lado los enfoques de "caja negra" del aprendizaje automático, tan de moda en los nuevos sistemas de "chat" lingüístico de IA abierta).

El análisis de sentimientos se basa en el procesamiento del lenguaje natural. Para cada texto de lenguaje natural (tras su 'limpieza' y normalización) se procesa un vector de sentimiento global, compuesto por múltiples párrafos. En cada párrafo se contabiliza el número de palabras asociadas a cada emoción básica, extraídas de un léxico ad hoc, obteniendo el porcentaje de palabras asociadas a cada emoción. El léxico es una lista de palabras inglesas (o de otros idiomas) con sus asociaciones a emociones básicas (a menudo se consideran ocho emociones básicas: ira, miedo, anticipación, confianza, sorpresa, tristeza, alegría y asco). Con cada palabra hay dos "sentimientos" asociados (de valencia negativa o positiva), o más valores en una gradación mayor (Mohammad, 2021). A continuación, a partir de los valores del vector de sentimientos -la valencia de los párrafos- y de la forma en que se distribuyen las emociones en el texto, se puede obtener una valoración estadística global. Además de este recuento emocional, aparece una trayectoria de valencia o sentimiento a lo largo del texto, que es un gráfico de la variación de la valencia emocional con respecto al tiempo narrativo. Se trata de un gráfico de trayectoria bastante revelador sobre la intencionalidad general y el estado de ánimo del texto.

A partir de aquí, buscando una interconexión fructífera de las metodologías de análisis de sentimientos con la inteligencia natural, lo que debería importar es la posibilidad de detectar no sólo las emociones directamente en colecciones de textos, la valencia global o el gráfico de trayectoria, sino destilar patrones contextuales que puedan desencadenar las distintas emociones. Forma parte del gran problema del "contexto", plagado de trampas semánticas para cualquier análisis automatizado. Los seis ejes de patrones que hemos identificado previamente, tres para casos interindividuales y otros tres para identidades colectivas, podrían abordarse provisionalmente mediante algunos de los nuevos métodos de *bootstrapping* y de clasificadores acoplados al aprendizaje profundo.

Para facilitar este tipo de exploración de la IA se necesita una postura parsimoniosa. Tal vez, del mismo modo que "los grupos dan forma a las emociones" y "las emociones dan forma a los grupos" (van Kleef and Fischer, 2016), podríamos afirmar que "los contextos dan forma a las emociones" y "las emociones dan forma a los contextos". Esta dialéctica puede llevarse a cabo mediante modelos de análisis de sentimientos dotados de léxicos adecuados en los que las palabras cargadas de emoción crean contextos (que se identifican como patrones), y la presencia de los patrones contextuales ayuda a reconocer más palabras emocionales. Los léxicos deben construirse con mayor complejidad interna, cada palabra proyectada inicialmente a todos los (tres o seis) ejes presentes, con una gradación de valencia más compleja, y estableciendo de algún modo un vínculo de refuerzo con las demás palabras emocionales presentes en la frase o el párrafo. Se trata de una metodología de *bootstrapping* que se diseñará y aplicará mediante PNL y análisis de sentimientos. Para empezar, podrían desarrollarse modelos más sencillos relacionados con la polaridad de cada eje del patrón.

En resumen, el rápido avance de estos nuevos campos del análisis de sentimientos, la minería de opiniones, la computación afectiva y la IA emocional, así como sus múltiples aplicaciones en los medios sociales, el marketing, la atención sanitaria, las encuestas sociales, la predicción política, etc., aunque bastante alejadas de algunos de los puntos de vista aquí defendidos, también podrían representar un grado de oportunidad para el avance de nuevos tipos de investigación sobre las emociones sociales. Que las futuras aplicaciones de la IA puedan trascender o no la actual "era del capitalismo de la vigilancia" (Zuboff, 2019), también puede depender de la relevancia de la investigación contrapuesta.

1.6. Aplicaciones del Análisis de Sentimientos

Como se ha comentado anteriormente, el análisis de sentimientos se trata de una disciplina que extrae las opiniones, sentimientos y pensamientos de las personas a partir de los datos de texto de los usuarios mediante métodos de Procesamiento del Lenguaje Natural (PLN) (Hasan et al., 2019). El análisis de sentimiento tiene muchas aplicaciones, que van desde el análisis de la opinión del cliente (Capuano et al., 2021; Pankaj et al., 2019; Păvăloaia et al., 2019), el análisis del estado de salud de pacientes (Tang et al., 2023; Weissman et al., 2019), hasta el análisis de crisis y desastres naturales (Mendon et al., 2021; Neppalli et al., 2017; Yuan et al., 2020). Además, avances tecnológicos como Blockchain, IoT, Cloud Computing y Big Data han ampliado el abanico de aplicaciones del Análisis de Sentimiento, permitiendo su uso en prácticamente cualquier disciplina (Wankhade et al., 2022). La Figura 2 muestra algunas de las aplicaciones más utilizadas en el análisis de sentimiento. Los dominios y sectores más significativos donde se aplica el Análisis de Sentimiento se describen a continuación.

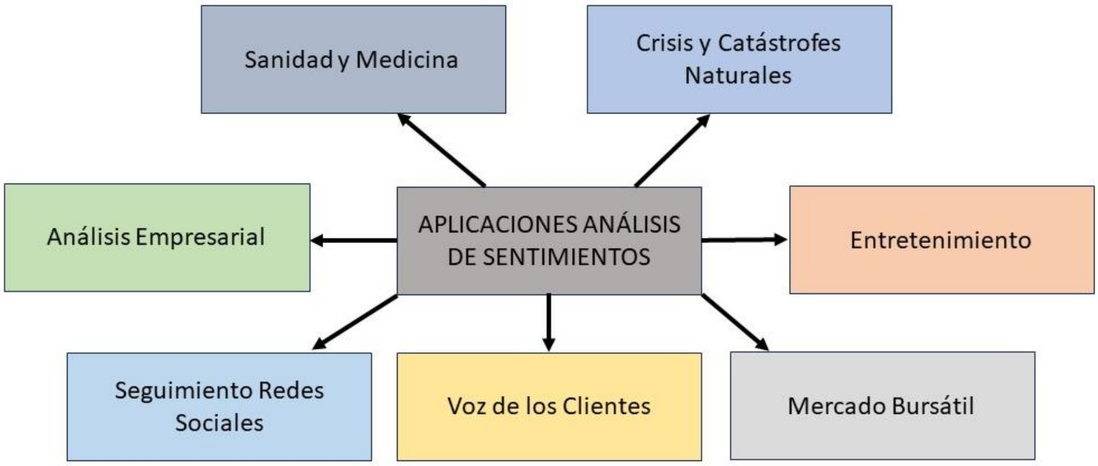

Figura 2: Aplicaciones del análisis de sentimientos. Se distinguen en la figura las diversas posibilidades existentes en los distintos sectores. Fuente: Elaboración propia.

1.6.1. Análisis empresarial

El análisis de sentimientos en el ámbito de la inteligencia empresarial ofrece numerosas ventajas (Wankhade et al., 2022). Por ejemplo, las empresas pueden utilizar los datos del análisis de sentimiento para mejorar los productos, investigar las opiniones de los clientes y desarrollar una estrategia de marketing innovadora. El uso más típico del análisis de sentimientos en el campo de la inteligencia empresarial es analizar las impresiones de los clientes sobre servicios o productos. Sin embargo, estos estudios no se limitan a los productores de productos; los consumidores pueden utilizarlos para revisar artículos y tomar decisiones más informadas. Las empresas, por su parte, pueden utilizar los resultados del análisis de sentimientos para introducir mejoras en los productos, examinar las opiniones de los consumidores o desarrollar un nuevo plan de marketing (Han et al., 2019). El análisis de sentimientos se utiliza con mayor frecuencia en la inteligencia empresarial para examinar las percepciones de los clientes sobre los productos o servicios. Durante ocho años, (Bose et al., 2020) el servicio de reseñas en amazon.com ha utilizado un léxico de emociones, que las clasifica en ocho emociones diferentes y dos estados de ánimo (positivo y negativo). Descubrieron que el análisis de sentimientos puede utilizarse eficazmente para identificar comportamientos y riesgos de los clientes y aumentar su satisfacción.

1.6.2. Sanidad y medicina

Este es uno de los sectores en los que se está utilizando el análisis de sentimientos en los últimos tiempos. Los datos pueden obtenerse de diversas fuentes, como encuestas, Twitter, blogs, artículos de noticias, reseñas, etc. (Carvalho and Plastino, 2021). Estos datos pueden analizarse para casos de uso, uno de los cuales es la evaluación de estándares y el análisis de novedades en el ámbito médico. Los expertos en la materia están investigando activamente para encontrar más usos del análisis del sentimiento y otras aplicaciones de PNL (Ebadi et al., 2021). Esta aplicación ayuda a los proveedores de servicios sanitarios a recopilar y evaluar el estado de ánimo de los pacientes, las epidemias, las reacciones adversas a los medicamentos y las enfermedades para mejorar la atención sanitaria en epidemias. En el trabajo de Jiménez-Zafraet y colaboradores (Jiménez-Zafra et al., 2019) se señalan las dificultades de aplicar el análisis de sentimientos en la atención sanitaria debido a las terminologías específicas y únicas utilizadas en el dominio. En el trabajo de Roccetti y colaboradores (Roccetti et al., 2017) se utilizaron mensajes de Twitter y facebook relativos a experiencias de pacientes como complemento para los análisis de la salud. A lo largo de un año, generaron aproximadamente cinco millones de tuits relacionados con el cáncer de mama, utilizando la API de streaming de Twitter. En conclusión, la aplicación general de estos modelos para analizar los datos generados por los pacientes en las redes sociales puede ayudar a mejorar la asistencia y determinar mejor las necesidades y opiniones de los pacientes (Wankhade et al., 2022).

1.6.3. Crisis y catástrofes naturales

En cuanto a las emociones y sentimientos que despiertan las crisis y desastres naturales, la percepción del peligro es esencial. Genera las reacciones emocionales más potentes (Barrett, 2006; Ekman, 1999; Russell and Barrett, 1999), en función de la proximidad a la zona de peligro y de otros factores de riesgo. Estas reacciones emocionales humanas tienen fuertes raíces evolutivas (Navarro and Marijuán, 2022). Pueden clasificarse en patrones de acción fijos, patrones de percepción-acción flexibles y patrones de percepción-acción superestructurales (Csany, 1988). Evaluar y monitorizar los sentimientos sobre las crisis y desastres naturales puede ser relevante para entender mejor esta coevolución y preparar mejor a la sociedad para futuros riesgos. El análisis de sentimientos o la minería de opinión de diferentes eventos y procesos se han explorado cada vez más (Clavel and Callejas, 2016). Los enfoques actuales de la minería de opiniones son múltiples y se centran en la clasificación de valencias positivas/negativas (análisis de sentimientos de polaridad (Chan and Liszka, 2013) y en la tipificación de los diferentes niveles de detalle que pueden aplicarse al análisis. Mediante la introducción de métodos estadísticos con una orientación más específica, pueden superarse algunos retos habituales en la minería de opiniones: (i) el conflicto inducido por el hecho de que una palabra de opinión pueda considerarse positiva o negativa en función de situaciones específicas; (ii) la probabilidad de que las personas no transmitan sus opiniones de la misma manera en circunstancias diferentes; (iii) la relativa ausencia de análisis de sentimientos en idiomas distintos del inglés o el chino (Vinodhini and Chandrasekaran, 2012), que son los idiomas dominantes, hasta la fecha, en estos estudios. Especialmente en los últimos años, se han publicado numerosos trabajos científicos que analizan el uso del análisis de sentimiento en Twitter para eventos muy diferentes: desastres naturales (Mendon et al., 2021), crisis de refugiados sirios (Öztürk and Ayvaz, 2018), referéndum Reino Unido-UE (Agarwal et al., 2018), impacto del Brexit (Ilyas et al., 2020), elecciones presidenciales y generales en diferentes países: EE. UU. (Caetano et al., 2018), Indonesia (Budiharto and Meiliana, 2018) e India (Sharma and Ghose, 2020), el mundial de fútbol (Patel y Passi, 2020) y el brote de COVID-19 (Garcia and Berton, 2021; H. Manguri et al., 2020; Patel and Passi, 2020). Los dos trabajos presentados en esta Tesis, sobre políticas del Covid 19 ((Turón et al., 2023) y la erupción del volcán Cumbre Vieja (Navarro et al., 2023), pueden ser incluidos también en esta recopilación.

1.6.4. Entretenimiento

El análisis de sentimientos se utiliza ampliamente en el ámbito del entretenimiento. Las críticas de películas, espectáculos y cortometrajes pueden analizarse para determinar la respuesta del espectador (Kumar et al., 2019). Esto no sólo ayuda a los espectadores a hacer una mejor elección, sino que también ayuda a que los contenidos de calidad ganen popularidad. El análisis de sentimiento a nivel de frase se ha utilizado comúnmente en este dominio para determinar el sentimiento general de las reseñas con mayor precisión. El sector del turismo y los viajes también ha intentado mejorar las experiencias de los clientes desarrollando aprendizaje automático y sistemas de recomendación de consumidores en línea basados en técnicas inteligentes de toma de decisiones basadas en datos (Jain et al., 2021a). Las decisiones de los consumidores se califican como positivas o negativas basándose en las reseñas en línea proporcionadas por los valiosos comentarios de los consumidores (Jain et al., 2021b).

1.6.5. Mercado bursátil

Otra de las aplicaciones del análisis de sentimientos es la predicción del precio de las acciones. Puede hacerse analizando todas las noticias sobre el mercado bursátil y prediciendo las tendencias de los precios de las acciones. Los datos pueden obtenerse de diversas fuentes, como Twitter, artículos de noticias, blogs, etc. El análisis de sentimiento a nivel de frase puede realizarse en estos textos, tras lo cual se decidirá la polaridad general de las noticias de la bolsa sobre una empresa concreta. En el trabajo de Xing y colaboradores (Xing et al., 2018) se utiliza para determinar si la tendencia será al alza o a la baja. Las noticias positivas tendían a llevar a una tendencia al alza, mientras que las noticias negativas tendían a llevar a una tendencia a la baja. Bitcoin y otras criptodivisas digitales se relacionan con la novedosa tecnología conocida como *Blockchain*. Los participantes de la red blockchain verifican las transacciones digitales utilizando métodos de consenso entre pares.

1.6.6. La voz de los clientes

Todas las opiniones de los usuarios procedentes de los centros de atención telefónica, correos electrónicos, encuestas, chats y web pueden combinarse para una evaluación conjunta. El análisis de sentimientos permitirá categorizar y organizar los datos para detectar tendencias y problemas y preocupaciones recurrentes (Kang and Park, 2014). El análisis de sentimientos puede ayudaren en la identificación de un grupo de clientes adecuado y el posterior desarrollo de la propuesta de valor, componentes ambos esenciales para el éxito de una operación comercial. Por otra parte, para estar al día y mantener el producto en demanda, debe seguirse continuamente el pulso de los clientes.

1.6.7. Seguimiento de las redes sociales

El análisis de sentimientos de los datos sociales permite monitorizar el sentimiento de los usuarios y clientes 24 horas al día, siete días de la semana, en tiempo real y cuando empiece a circular algo desagradable. Ello permite rápidamente responder y reforzar la imagen, tanto en imprevistos negativos como cuando se reciban menciones favorables (Bohlouli et al., 2015). También se obtiene información fiable sobre los usuarios y clientes, que puede servir para comparar la evolución de una temporada a otra y para el proceso de toma de decisiones. Dado que los particulares aportan en general sus comentarios desde los puntos de vista más honestos sobre productos, servicios, empresas y entidades, es importante estar al tanto de esta imagen de cara al resto del público.

2. METODOLOGÍA E HIPÓTESIS DE TRABAJO

2. METODOLOGÍA E HIPÓTESIS DE TRABAJO

En cuanto a la metodología que se seguirá en la presente investigación, consta de dos aspectos aparentemente separados pero claramente interrelacionados. Por un lado, el estudio de las redes sociales naturales y las emociones predominantes en dichas redes, teniendo en cuenta enfoques teóricos, los desarrollos ya realizados alrededor del sociotipo y los diseños de trabajo de campo (p.ej., en el mencionado trabajo *"The cost of loneliness"* (Navarro et al., 2022). Es decir, las herramientas sencillas. Y por otro lado, están los métodos de *machine learning* e inteligencia artificial aplicados al análisis de emociones (*sentiment analysis*) en mensajes intercambiados en redes sociales o en noticias de prensa. Es decir, las herramientas complejas.

En el desarrollo de estas ideas se han formulado las siguientes hipótesis de trabajo:

- **El sociotipo**: a partir de la acción de nuestro cerebro social surge una necesidad real de vinculación social que tiene raíces evolutivas ("proviene de nuestros genes"). El sociotipo representa nuestra socialidad adaptativa, el conjunto relacional compuesto por los círculos vinculares de la familia y los parientes, los amigos y los conocidos. La crisis del sociotipo de la persona la conocemos como *"soledad no deseada"*. Detectarla mediante herramientas sencillas resultaría plausible a partir del marco conceptual del sociotipo.

- **Las emociones sociales**: necesitamos reacciones instintivas para lograr nuestra adecuación individual en el grupo y para construir y mantener nuestro propio sociotipo, así como para desarrollar la "inteligencia social" colectiva en el grupo. Frente a las seis emociones fundamentales de Paul Ekman (ira, alegría, asco, tristeza, sorpresa, miedo) podríamos encontrar otras (p.ej., admiración, envidia, superioridad, inferioridad, envidia, generosidad, aprecio, resentimiento, amor, odio, estrés, relajación, etc.) que estarían mucho más cerca de las experiencias cotidianas en los grupos sociales.

- **La toma de decisiones colectivas,** que va acompañada o precedida, o seguida *a posteriori*, por toda una serie de intercambios a través de las redes sociales en los que resultaría plausible detectar analíticamente la presencia y la dominancia de determinadas emociones netamente sociales.

- **Las herramientas de análisis de sentimientos,** como las técnicas de Inteligencia Artificial y *machine learning*, que resultan adecuadas para detectar y evaluar el estado anímico individual y colectivo.

3. OBJETIVOS

3. OBJETIVOS

3.1. Objetivo principal

El objetivo principal de la tesis consiste en establecer una nueva comprensión de las dinámicas emocionales subyacentes a las relaciones sociales y a la toma de decisiones, así como poder aquilatar su impacto en diferentes ámbitos (economía, salud, cultura…). Es importante destacar que, en lugar de considerar sólo las emociones básicas convencionales, ponemos el foco en aquellas otras que, aun siendo de menor potencia expresiva, tienen una mayor presencia en las relaciones sociales cotidianas. Las podríamos denominar "emociones sociales". Al respecto, se necesita establecer un sistema de referencia para la organización y expresión de dichas emociones, las cuales, reiteramos, son las que se manifiestan en nuestro sociotipo y en la estabilidad del propio grupo social. Para categorizarlas, se deben utilizar conceptos auxiliares como valencia, excitación, efectos de vinculación social y, quizás, ganancias-costes de adaptabilidad (*fitness*), junto con el conocido "cuadro cartesiano" de clasificación de las emociones basado en la valencia y la excitación (Posner et al., 2005).

Al respecto se han establecido los siguientes Objetivos Específicos (OE):

3.2. OE1: Análisis de la falta de relaciones sociales

- Desarrollar y validar una versión abreviada del Cuestionario de Sociotipo (SOCQ), para evaluar las relaciones sociales de la población general, dirigida específicamente a la población que con más frecuencia sufre la soledad no deseada: los mayores.

- Construir una escala dicotómica de 4 ítems (SOCG-4) a partir de los 12 ítems de la escala original del SOCQ, de forma que pueda servir para discriminar entre los pacientes de atención primaria y de la clínica geriátrica, ayudando al facultativo a identificar a aquellos que necesitan atención social o intervención psicosocial.

- Realizar una prueba o estudio piloto para evaluar las relaciones sociales en el uso de la clínica geriátrica.

3.3. OE2: Análisis de Sentimientos

- Desarrollar un conjunto de herramientas para el análisis de sentimientos en las comunicaciones escritas que se vierten en las redes sociales (se considerará *Twitter* esencialmente en el presente estadio del proyecto), y también en las noticias publicadas en la prensa.

- Construir un Lexicón en español ad hoc para el proyecto. El lexicón contendrá un listado de palabras en español y sus asociaciones con ocho emociones básicas (ira, miedo, anticipación, confianza, sorpresa, tristeza, alegría y asco) y dos sentimientos (negativo y positivo).
- Extraer indicadores relevantes para el problema a partir de minería de opiniones en los mensajes emitidos en las redes sociales y prensa escrita.
- Analizar la estructura de la narración a lo largo del texto obteniendo gráficos de trayectorias, es decir, una representación del tiempo narrativo frente a la valencia emocional.
- Analizar la distribución de las emociones a lo largo del texto utilizando estadísticos descriptivos y realizar análisis gráficos de los datos en busca de patrones de comportamiento. Extraer y difundir el conocimiento generado. Visualización gráfica.

3.4. OE3: Evaluar mediante herramientas de Inteligencia Artificial (análisis de sentimientos y machine learning) el estado anímico individual y colectivo.

- Desarrollar un software/algoritmo que permita la búsqueda y extracción automática de textos en las diferentes redes sociales y medios escritos.
- Desarrollar herramientas de Inteligencia Artificial y *machine learning* para identificar el estado anímico (individual y colectivo) a través del análisis de textos escritos.

3.5. OE4: Aplicaciones sociosanitarias. Estado anímico y aplicación al COVID-19

- Evaluar mediante Inteligencia Artificial y *machine learning* el estado actual de opinión colectiva de la ciudadanía en relación con el ámbito sociosanitario.
- Evaluar mediante dichas técnicas el impacto en la mentalidad social de la pandemia COVID-19.

- Evaluar el efecto de las distintas decisiones y políticas públicas sociosanitarias en la opinión de la ciudadanía respecto a la pandemia COVID-19.

3.6. OE5: Aplicaciones a otras crisis y catástrofes naturales

- Evaluar mediante Inteligencia Artificial y *machine learning* las reacciones sociales provocadas por el impacto de la actividad del volcán del Cumbre Vieja (Islas Canarias) en 2021.
- Clasificar las noticias de prensa dentro de un determinado episodio o periodo de tiempo durante el desarrollo o evolución de una catástrofe natural.

4. CARACTER INNOVADOR

4. CARACTER INNOVADOR

1) La medición de las relaciones sociales de las personas mayores es un importante tema de investigación sociológica (y sociométrica), así como una profunda preocupación biomédica, especialmente después de la pandemia COVID-19. Esta medición es esencial para los facultativos de atención primaria, las consultas geriátricas y los servicios de atención social.

2) En el análisis de los patrones emocionales colectivos en las redes sociales, se combinan técnicas de *machine learning* e Inteligencia Artificial con técnicas de estadística multivariante para evaluar las comunicaciones de forma masiva, detectando el cambiante estado de ánimo de los ciudadanos, en particular durante la pandemia COVID-19, como respuesta a las decisiones políticas adoptadas por los diferentes niveles de la Administración.

3) Todo ello se realiza a partir del análisis de un importante volumen de textos escritos (*Twitter*, pero igualmente se puede realizar con textos procedentes de la prensa, internet, redes sociales...). El análisis de sentimientos o "minería de opinión" que se propone (Salazar et al., 2008) permite el análisis cuantitativo del texto mediante la extracción de información subjetiva a partir del examen de la polaridad, es decir, la connotación positiva o negativa del lenguaje utilizado y el tipo de "arco semántico" empleado en cada texto (Salazar et al., 2010). Este tipo de análisis ayudará a evaluar el estado de ánimo de la población e impulsar el apoyo social a las políticas orientadas a mitigar los efectos de la pandemia.

4) En su aplicación a la Teoría de Decisiones, la novedad se encuentra en cómo integrar metodológica y tecnológicamente el tratamiento de datos numéricos que reflejan las preferencias del decisor, con contenidos en texto que reflejan los argumentos que soportan las diferentes posiciones y decisiones. Para ello se combinarán técnicas de Inteligencia Artificial y *machine learning* con procedimientos estadísticos, y si es preciso, también cabe aplicar herramientas criptográficas apropiadas para garantizar las propiedades de seguridad exigidas a los procesos de e-participación (Toncovich et al., 2013).

5) Las aplicaciones prácticas abordan dos problemas de indudable relevancia y actualidad: la soledad en la tercera edad y la gestión de las pandemias y desastres naturales.

5. TRABAJOS PRESENTADOS

5. TRABAJOS PRESENTADOS

A. *THE COST OF LONELINESS: ASSESSING THE SOCIAL RELATIONSHIPS OF THE ELDERLY VIA AN ABBREVIATED SOCIOTYPE QUESTIONNAIRE FOR INSIDE AND OUTSIDE THE CLINIC*

Este primer trabajo analiza la ausencia de relaciones sociales en las personas mayores, un importante tema de investigación sociométrico y una profunda preocupación biomédica, así como económica y social, especialmente tras el COVID-19. Este artículo desarrolla un amplio trabajo de campo a partir del que se obtiene una escala dicotómica de 4 ítems (SOCG-4), desde los 12 ítems de la escala original del SOCQ, de forma que pueda servir para discriminar entre los pacientes de atención primaria y de la consulta geriátrica. La escala de 4 ítems del Sociotipo Geriátrico (SOCG-4) aparece como un instrumento de medida válido para su uso en la clínica y en otras instancias de atención social.

B. *NATURAL INTELLIGENCE AND THE 'ECONOMY' OF SOCIAL EMOTIONS:* A CONNECTION WITH AI SENTIMENT ANALYSIS

En este trabajo se aborda el concepto de Inteligencia Natural como un nuevo camino en una variedad de problemas teóricos y aplicados sobre las emociones sociales. Aquí proponemos una vía diferente, centrada en el avance del ciclo vital. El ciclo vital en su integridad se convierte en el núcleo de los procesos informativos de la inteligencia natural, incluyendo la expresión coherente de las emociones a lo largo de las ocasiones de maximización de la aptitud. En las sociedades humanas, la "economía" global de las emociones sociales se manifiesta, mostrándose en la conspicua interacción entre los procesos de vinculación y las diferentes clases de emociones sociales. El vínculo esencial entre la inteligencia natural, las emociones y el ciclo vital de los individuos puede armonizar con los actuales avances -y puntos ciegos- de campos de la inteligencia artificial como el "análisis de sentimientos".

C. *EVOLUTION OF SOCIAL MOOD IN SPAIN THROUGHOUT THE COVID-19 VACCINATION PROCESS: A MACHINE LEARNING APPROACH TO TWEETS ANALYSIS*

Este trabajo presenta un nuevo enfoque basado en la combinación de técnicas de aprendizaje automático, en particular, el análisis de sentimientos mediante léxicos, y métodos estadísticos multivariantes para evaluar la evolución del estado social de ánimo a lo largo del proceso de vacunación COVID-19 en España. Se han analizado 41.669 tuits españoles publicados entre el 27 de febrero de 2020 y el 31 de diciembre de 2021, se evaluaron diferentes sentimientos mediante una lista de palabras en español y sus asociaciones con ocho emociones básicas (ira, miedo, anticipación, confianza, sorpresa, tristeza, alegría y asco) y tres valencias (neutra, negativa y positiva). Los resultados obtenidos son muy ilustrativos de los diferentes cambios de opinión emergentes, midiendo el estado de ánimo de los ciudadanos a través de la valencia colectiva, y detectando la prevalencia de diferentes emociones en las sucesivas etapas del proceso de vacunación. La presente combinación en modelos formales de información objetiva y subjetiva proporciona una visión más precisa de la realidad social, en este caso del proceso de vacunación de COVID-19 en España, lo que permitiría una resolución más eficaz de los problemas.

D. *PRESS MEDIA IMPACT OF THE CUMBRE VIEJA VOLCANO ACTIVITY IN THE ISLAND OF LA PALMA (CANARY ISLANDS): A MACHINE LEARNING AND SENTIMENT ANALYSIS OF THE NEWS PUBLISHED DURING THE VOLCANIC ERUPTION OF 2021*

En este trabajo hemos utilizado como fuente de información una extensa muestra de artículos de prensa (en contraposición a la inmensa mayoría de los artículos científicos, que usan las redes sociales) sobre la erupción del volcán Cumbre Vieja. Hemos mostrado cómo las emociones y sentimientos expresados en los medios de prensa pueden ser analizados eficientemente mediante técnicas de IA para evaluar mejor el impacto social y económico de una catástrofe en el momento en que se produce. Este enfoque permitiría mejorar el diseño de las herramientas para estudiar la evolución de las repercusiones sociales, económicas y medioambientales de un fenómeno natural y evaluar cómo cambian con el tiempo y las decisiones tomadas. Asimismo, los resultados de este estudio sobre la prensa pueden contribuir a compensar posibles sesgos en la difusión de información durante las crisis y catástrofes naturales y a mejorar el análisis de las estrategias seguidas para reducir su impacto.

6. COPIA DE LAS PUBLICACIONES INCLUIDAS

International Journal of
Environmental Research and Public Health

MDPI

Article

The Cost of Loneliness: Assessing the Social Relationships of the Elderly via an Abbreviated Sociotype Questionnaire for inside and outside the Clinic

Jorge Navarro [1,*], Miguel Cañete [2], Francisco J. Olivera [3], Marta Gil-Lacruz [4], Ana Gil-Lacruz [5] and Pedro C. Marijuán [6]

1 Grupo de Decisión Multicriterio Zaragoza (GDMZ), Faculty of Economics, University of Zaragoza, 50005 Zaragoza, Spain
2 Psychology and Sociology Department, Education Faculty, University of Zaragoza, 50009 Zaragoza, Spain; mcanete@unizar.es
3 Department of Psychiatry, Hospital San Jorge (Psychogeriatric Program), 22004 Huesca, Spain; joliverap@aragon.es
4 Psychology and Sociology Department, Health Science Faculty, University of Zaragoza, 50009 Zaragoza, Spain; mglacruz@unizar.es
5 Business Department, School of Engineering and Architecture, University of Zaragoza, 50018 Zaragoza, Spain; anagil@unizar.es
6 Independent Scholar Affiliated to Bioinformation Group, Instituto Aragonés de Ciencias de la Salud (IACS), 50009 Zaragoza, Spain; pcmarijuan.iacs@aragon.es
* Correspondence: jnavarrol@unizar.es

check for updates

Citation: Navarro, J.; Cañete, M.; Olivera, F.J.; Gil-Lacruz, M.; Gil-Lacruz, A.; Marijuán, P.C. The Cost of Loneliness: Assessing the Social Relationships of the Elderly via an Abbreviated Sociotype Questionnaire for inside and outside the Clinic. *Int. J. Environ. Res. Public Health* **2022**, *19*, 1253. https://doi.org/10.3390/ijerph19031253

Academic Editor: Ricardo Pagan

Received: 26 November 2021
Accepted: 21 January 2022
Published: 23 January 2022

Abstract: Gauging the social relationships of the elderly is a significant sociometric research subject and a deep biomedical concern—particularly after the COVID-19 pandemic. It is imperative for facultatives in primary care, for geriatric clinics, and for social care services. In this respect, this article explores the validity of an abbreviated version of the Sociotype Questionnaire (SOCQ), a tool previously developed by the authors for assessing the social relationships of the general population, now specifically addressed to the elderly population. The aim is to construct a 4-item dichotomous scale (SOCG-4) out of the 12 items of the original scale of the SOCQ, so that it can serve to discriminate among the patients in primary care and the geriatric clinic, helping the facultative to find those in need of social care or of psychosocial intervention. The population data have been obtained from a series of previous studies on social relationships in different segments of the elderly population (Ntotal = 915). The resulting abbreviated version of SOCG-4 was extracted by means of confirmatory factor analysis, with the congruence, validity, and relationship with the determinants as close to optimal. The significant correlations with SOCQ (0.82), UCLA (−0.55), Barthel (0.40), and other relevant tests are obtained. The test was also put to trial in a pilot study, being applied to 150 subjects via phone surveys, home visiting, and geriatric clinic—it becomes particularly useful for assessing the social relationships in geriatric clinic use. The 4-item Geriatric Sociotype scale (SOCG-4) appears as a valid measurement instrument for use in the clinic and in other social care instances.

Keywords: loneliness; social relationships; sociotype questionnaire; elderly isolation risk; geriatric clinic

1. Introduction

The metrics of social relationships of the elderly acquired, even before the COVID-19 pandemic, an increasing importance due to two major trends in contemporary societies. On the one hand, there is the general improvement of public health and the parallel demographic transition, which are causing a remarkable increase in the older population with inevitable tensions in health systems and in social care services. On the other hand, there are profound cultural changes related to lifestyle, communication technologies, increased individualism, diminished "social capital," and weakened social-bonding structures, which remarkably diminished the preferred relational activities (face-to-face talk) and increased feelings of personal isolation. The impact on physical health and mental health of this rising social condition of isolation is difficult to estimate, as well as the burden it represents for social care and public health systems [1]. Some authors claim there is a genuine "epidemic of loneliness" to confront [2–5]. Without a doubt, this worsening situation makes it necessary to gather more kinds of quantitative data concerning the different contexts of analysis, intervention, and clinics among the elders. In this sense, developing a very brief metrics on social relationships (and particularly the lack of this), to allow the straightforward use for geriatric clinics, is the *leit motif* of this paper.

Since the COVID-19 pandemic, loneliness has become an even more significant concern [6]. The pandemic limited social contact and promoted teleworking and telematic relationships. The World Health Organization (WHO) already confirmed that the rate of loneliness has risen, and the prevalence of mood disorders has increased (above all, depression and anxiety, and also self-harm and suicide) and exacerbated pre-existing mental health conditions [6]. Economic analysis is throwing light on the multiple costs attributable to, or resulting from, the mental health burden accompanying the pandemic [7]. Loneliness and its associated mental disorders generate substantial costs for the individual, family, employer, and community, payments ranging from medical bills to sick leave days, to medication expenses, to home care services, to nursing homes.

Overall, the direct impact of loneliness on mental and physical health, as well as the social burden it causes, has already been highlighted by a number of researchers in fields as diverse as social economics [8–12], social psychology [13–15], mental health and psychiatry [2,16–18], and physiology [19–21], not to speak of medical practitioners directly involved in primary care, family care, and geriatric clinics [3,4,22–25]. One of the main concerns is that the increase in the aging population and growing loneliness are putting the sustainability of public healthcare systems at stake [26]. Indeed, the research on aging has come into a new era that marks an inflection point, with unique medical, economic, and societal implications [21]. Additionally, all these preexisting problems were magnified by the social impact of the COVID-19 pandemic.

Thereafter, gauging the level of the social interaction of the elderly population becomes an important research goal. In this respect, the Sociotype Questionnaire (SOCQ) already developed by this team [27] was addressed to evaluate the main dimensions of social interaction for the general population. Based on the "Social Brain Hypothesis" [28], it emphasized the adaptive nature of our social relationships and the general convergence on widely shared classes of social bonds and related face-to-face exchanges [29]. Although this Sociotype Questionnaire (counting only 12 items) was also applicable to the elderly, it was not sufficiently agile for use in the geriatric clinic, in primary care, or in social care screening.

The challenge, regarding a variety of geriatric interventions and in primary care, is the creation of an abbreviated questionnaire of preferably 4 items, reminiscent of the blocks of the Mini-Mental State Examination (MMSE) and other reduced questionnaires, specifically tailored to the necessities of the geriatric clinic, and covering the main dimensions of sociability. Thus, the objective of the present paper is to construct a 4-item dichotomous scale (SOCG-4) out of the 12 items of the original scale of the SOCQ, so that it can serve to discriminate among the geriatric patients in the clinic, helping the facultative to find those in need of social care or of psychosocial intervention. The reduced number of items, 4, and the dichotomic nature of the responses should configure an easy instrument to be kept handy on the desk of the geriatric or primary care facultative. In fact, there is a dearth of adequate indicators for systematic use in the geriatric clinic, so that the practitioner might easily detect the condition of loneliness [30,31]. Contributing to fill in that detection gap is an essential aspect of this research. Needless to say, the authors are fully aware that no single indicator of perceived or actual social isolation can fully assess the degree to which an individual is lacking social resources [31].

In addition to the obtention of the abbreviated questionnaire, we included a brief pilot study in Appendix A, where SOCG-4 was applied to a total of 150 subjects in 3 different modalities: via phone surveys, home visiting, and the geriatric clinic (50 subjects each). The SOCG-4 appears as a valid instrument, particularly useful in the geriatric clinic, and also in the other social care instances for which an easy discriminant tool may be needed.

2. Materials and Methods

The data of four previous studies were integrated to obtain the abbreviated Sociotype Geriatric Questionnaire. These studies were part of an official research project granted by the Spanish Ministry of Health (FIS PI12/01480): "The Sociotype: A new metrics on the structure and dynamics of social relationships." The project incorporated different studies, some of them exclusively concerning the elderly population, which were addressed towards different research purposes, namely the assessment of social networking, communication modalities, institutionalization versus home dwelling, and gender differences in socialization. These studies shared basic concepts, methodologies, and survey platforms. All of them were quantitative, though initially a qualitative study was also undertaken [27].

2.1. Component Studies

2.1.1. Study 1

The total sample for this study comprised of 1075 participants from the general population (18–95 years). Of these, 208 were 65 years or older and were incorporated into the SOCG-4 (N1). An online internet platform, "SurveyMonkey®", was used for data gathering and the statistical support. The questions included a range of sociodemographic variables, the entire Sociotype Survey, and several complementary tests on loneliness, general health, and personality, including the Revised UCLA Loneliness Scale (RULS), the General Health Questionnaire (GHQ-12), and the Eysenck Personality Questionnaire-Revised (EPQ-R). This series of questionnaires was self-administered, with the availability of support staff when needed. As a result of the study, a 16-item Sociotype Questionnaire, SOCQ, was validated [27], which was counted using 4 main relational dimensions: family, friendships, acquaintances, and work/study when applicable. Otherwise, when the latter dimension was not applicable, i.e., for the elderly, SOCQ became a 12-item questionnaire.

2.1.2. Study 2

This geriatric study involved a total of 275 participants (65 years or older). It was based on the same procedures and questionnaires as Study 1 (Sociotype Survey, UCLA, GHQ-12, and Eysenck). It addressed the elderly population living in their own homes. The goal of this study was to observe the quantitative differences of the elderly versus other age segments, regarding the averages of both social networking and daily talking time [29].

2.1.3. Study 3

This geriatric study involved 283 participants (65 years or older), most of them enrolled in active and healthy aging activities organized by the Social Care Regional Institute (IASS), which oversees the active and healthy aging programs in the Spanish region of Aragon. Apart from the Sociotype Survey, this included several questionnaires that evaluated the specific characteristics of the elderly: the geriatric depression scale, the Pfeiffer questionnaire, and the Barthel ADL index. Its aim was to provide a comparison between age segments within the elderly population and to investigate the "minimum daily talking" (Navarro et al., in preparation).

2.1.4. Study 4

This geriatric study compared physical and mental health status, and assessed the social relationships od elderly people in two different circumstances: institutionalized in residences versus living at home. It included the Sociotype Survey and some other questionnaires that evaluated in detail the mental and physical characteristics of the elderly:

the geriatric depression scale, Pfeiffer questionnaire, Barthel ADL index, MiniMental, IAE cumulative index, and EuroQol-5D. The study population involved 200 individuals (70–97 years old), for which 100 individuals were institutionalized in residences (private and public) and 100 were living at home and attended daycare centers. All of them were interviewed by a psycho-geriatric facultative.

2.2. Participants

The total number of participants included in the present work is N = 915 ($N1$ = 208, $N2$ = 275, $N3$ = 283, and $N4$ = 200, respectively, belonging to studies 1, 2, 3, and 4. The inclusion criteria were being 65 years or older (70 years old in Study 4), being able to read and write Spanish, and not suffering from severe psychosis or not being diagnosed with severe cognitive impairment. The different samples were mainly composed of Caucasian adults, between 65–97 years old (mean = 76.95; SD = 8.02). All of them were Spanish (with diverse regional backgrounds, but mostly from Aragon); 70.6% were women and 29.4% were men, between the ages of 65–97 years (mean = 76.95; SD = 8.02); 40.7% of them had a partner or were married and 44.7% were widows/widowers; 38.6% lived alone; 33.9% lived with their partner; 11.5% lived in residences; and 16.5% accounted for other conditions. The main sociodemographic characteristics of all the participants are shown in Table 1.

Table 1. Sociodemographic data.

		Number	**% Population**
Gender	Women	646	70.6%
	Men	269	29.4%
	Total	915	100.0%
Age segments	From 65 to 75	333	38.5%
	From 76 to 85	406	46.9%
	More than 85	126	14.6%
	Total	865	94.5%
Stable relationships	Married/with partner	372	40.7%
	Single	102	11.1%
	Separate/divorced	32	3.5%
	Widow/widower	409	44.7%
	Total	915	100.0%
Connivence	Alone	353	38.6%
	Partner	310	33.9%
	Partner and offspring	50	5.5%
	Other family	53	5.8%
	Friends	2	0.2%
	Residence	105	11.5%
	Other	42	4.6%
	Total	915	100.0%

Table 1. *Cont.*

		Number	% Population
Education	Illiterate	28	3.1%
	No studies	372	40.7%
	Primary	324	35.4%
	High school	99	10.8%
	University	79	8.6%
	Other	13	1.4%
	Total	915	100.0%
Income	No income	35	7.3%
	Minimal pension	137	28.5%
	Average pension	232	48.3%
	High pension	61	12.7%
	Maximal pension	15	3.1%
	Total	480	52.5%

2.3. Procedure and Ethics

The procedures for Study 1 and Study 2 took approximately thirty minutes. Each participant was given information about the study, which included the aims of the project, the advantages/disadvantages of participation, a letter of informed consent, and an assurance of anonymity (in line with the Spanish Organic Law 15/99 on the Protection of Personal Data, and Law 41/02 on Patient Autonomy). A research psychologist or a hospital nurse were on hand to give support where required. The survey of Study 3 implied a similar procedure, though the response time was shorter, in the order of twenty minutes. The methodologies and questionnaires of these studies were previously approved by the Ethical Committee of Aragón (CEICA), Spain (Act: CP13/2014). In the case of Study 4, the response time was in the order of fifty minutes, and its approval was granted by the Ethical Committee of the Hospital Complex of Navarre (Act: 18 January 2017).

2.4. Statistical Procedure

The statistical analysis was conducted with SPSS software (IBM SPSS Statistics for Windows, Version 24.0, IBM, Armonk, New York, NY, USA). The description of the population characteristics was by means of numbers and percentages for the categorical variables, whilst the mean and standard deviations (SDs) were employed for the quantitative variables.

The objective of the statistical analysis was to construct a 4-item dichotomous scale (SOCG-4) out of the 12 items of the original scale of the SOCQ. To achieve that goal, three phases were established:

- First phase: selection process of the most representative item within each of the three dimensions of the SOCQ. To accomplish this, the weights from the exploratory factor analysis by the Principal Component method were taken as references (previously making sure that these factors for elderly people were no different from those obtained for the general population).
- Second phase: for the selection of a 4th item, we searched among all the other discarded items (removing those previously selected), and we selected that which obtained the highest correlation with the GDS, Barthel, Pfeiffer, EPQ, UCLA, GHQ12, MMSE, EuroQol-5D, and Goldberg scales (taking into account that each test only applied to some determined population).

- Third phase: once we selected the 4 items, they were dichotomized assigning the value of 1 to the answers of the upper categories of the SOCQ and the value 0 to the remaining categories. Then, the validity of this new, "binary" scale was confirmed.

In order to confirm the validity of the procedure, the correlations of the resulting scale (SOCG-4) were recalculated with all the tests referred to in the second phase, corroborating that these correlations were significant, and that the questionnaire had predictive validity.

3. Results

Given the methodology proposed, which in its first phase involved selecting the weights from the exploratory factor analysis by the principal component method, it was necessary to previously inspect the weights and h^2 values corresponding to the SOCQ questions for the aging population versus the general population. This could be conducted via the data of Study 1, where the values for general population were obtained, but a fraction of 208 individuals corresponded to the aging population.

3.1. SOCQ Differential Results in Study 1: The Aging Population versus General Population

By searching for the reliability of the SOCQ questions among the aging population, we can observe, in Table 2, that the results obtained also justify the 12 questions for the aging population; however, we observe an interesting difference regarding the weights and the h^2 values of most of those questions with respect to general population.

Table 2. Psychometric features of the SOCQ * (for the aging population in Study 1, *N*1 = 208).

SOCQ *	Mn	SD	DI	h^2	w_1	w_2	w_3
Family							
1. I speak and relate with my family	4.35	1.13	0.48	0.79	0.87	0.07	0.15
2. My family is important to me	4.73	0.85	0.40	0.66	0.80	0.03	0.11
3. The family members care about me	4.40	1.18	0.46	0.78	0.87	0.08	0.11
4. I have fun and laugh with my family	3.63	1.40	0.48	0.62	0.72	0.32	0.04
Friends							
5. I speak and relate with my friends	3.21	1.71	0.69	0.85	0.10	0.90	0.17
6. I have friends to tell and share problems	2.89	1.85	0.65	0.82	0.05	0.88	0.21
7. I consider it important to maintain relationships with friends	3.78	1.62	0.63	0.76	0.11	0.85	0.16
8. I have fun and laugh with my friends	3.02	1.64	0.69	0.77	0.23	0.83	0.18
Acquaintances							
9. I speak and relate comfortably with acquaintances	3.83	1.27	0.57	0.54	0.27	0.29	0.62
10. It is difficult for me to make conversation with people I do not know (r)	3.08	1.59	0.32	0.59	0.04	0.05	0.76
11. It is easy for me to win support from acquaintances	2.57	1.58	0.29	0.31	0.02	0.19	0.53
12. Relations with my acquaintances are forced (r)	3.76	1.27	0.41	0.60	0.14	0.12	0.75
% of variance (real-data)					17.73	38.59	11.12

SOCQ * Exploratory Factor Analysis; Mn = mean. SD = standard deviation; w_1, w_2 and w_3 = weights on the first-order factors; h^2 = communality; and r = reverse score. For all items from 1 to 12, the range of values is 0–5.

As Table 2 clearly shows that the new weights upon which the present study is based are different from those in Table 3 below, that reproduces the values corresponding to the general population [27]. Following the methodology proposed, the items selected in Phase 1 were different in at least two cases (the 2nd and 3rd, questions, and probably the 4th one too—see later). Subsequently, the differences found between the general population and the elderly become the determinants for the extraction of the abbreviated scale we are looking for.

Table 3. Psychometric features of the SOCQ (for the general population, N = 1075).

General SOCQ	Mn	SD	h^2	w_1	w_2	w_3
Family						
1. I speak and relate with my family	4.39	0.97	0.81	−0.16	0.94	0.02
2. My family is important to me	4.74	0.76	0.83	−0.12	0.91	0.10
3. The family members care about me	4.49	1.00	0.64	−0.04	0.81	−0.01
4. I have fun and laugh with my family	3.65	1.20	0.43	0.26	0.55	−0.12
Friends						
5. I speak and relate with my friends	3.44	1.48	0.81	0.89	−0.06	0.09
6. I have friends to tell and share problems	3.45	1.65	0.83	0.92	−0.07	0.06
7. I consider it important to maintain relationships with friends	4.14	1.39	0.81	0.90	−0.01	0.01
8. I have fun and laugh with my friends	3.59	1.41	0.68	0.82	0.09	−0.11
Acquaintances						
9. I speak and relate comfortably with acquaintances	3.61	1.19	0.47	0.06	0.12	0.61
10. It is difficult for me to make conversation with people I do not know (r)	3.19	1.33	0.34	−0.01	−0.08	0.61
11. It is easy for me to win support from acquaintances	2.29	1.48	0.24	0.08	−0.09	0.52
12. Relations with my acquaintances are forced (r)	3.53	1.05	0.42	−0.02	0.05	0.63
% of variance (real-data)				38.70	18.80	13.90

SOCQ Exploratory Factor Analysis; Mn = mean; SD = standard deviation; w_1, w_2 and w_3 = weights on the first-order factors; h^2 = communality; and r = reverse score. For all items from 1 to 12, the range of values is 0–5.

3.2. Construction of the Abbreviated Geriatric Sociotype Scale (SOCG-4)

Phase 1: following the factorial analysis based on the principal components method, we obtained a Kaiser–Meyer–Olkin index of the sampling adequacy of 0.845, with the Bartlett sphericity test $\chi 266$ = 5448.963 and $p < 0.001$. These values indicate that the factorial model is good and that it is therefore appropriate to carry out this technique based on the correlations found between the items. We used the Varimax orthogonal rotation procedure since the SOCQ was composed of three independent dimensions [27].

After conducting the factorial analysis, we found a 67.44% of total variance explained by the 3 factors, the most important being the friends factor with an explained variance of 38.59%, followed by the family factor with 17.73%, and finally the acquaintances factor with 11.12%. Since our interest was that the three dimensions of the SOCQ were represented in the final scale, we incorporated an item from each of the three dimensions—the one with the highest factorial weight (see Table 2). The selected items were: the 1st item in the family dimension (w_1 = 0.873), the 5th item in the friends dimension (w_2 = 0.902), and the 10th item in the acquaintance dimension (w_3 = 0.765).

Phase 2: we discarded for this phase the three items already selected. Additionally, for the selection of the fourth item of the simplified scale, we observed the correlations with the remaining scales used (GDS, Barthel, Pfeiffer, EPQ, UCLA, GHQ12, MMSE, EuroQol-5D, and Goldberg), considering the different populations associated. Of the 9 remaining items of the SOCQ, the one that had a greater correlation with these scales was the 8th item, belonging to the friends dimension. Therefore, this was chosen as the 4th item of the abbreviated scale. The final result was, then, the following scale:

1. *I talk and relate with my family*;
2. *I talk and relate with my friends*;
3. *(R)It is difficult for me to make conversation with people I do not know*;
4. *I have fun and laugh with my friends*.

Phase 3: the last step of the procedure was to dichotomize the four items and check the psychometric properties of the dichotomized final scale. The three highest values of response in each of these 4 items were assigned the value of 1, and the three lower ones the value of 0. The reliability coefficient (Cronbach's α) for the final scale was 0.57 (N = 915) and the correlations with the remaining tests are shown in Table 4.

Table 4. Correlations of the dichotomous simplified SOCG-4 Scale.

		Range of Values	Mn	SD	SOCG-4 Correlation	*N*
SOCG-4		0–4	2.98	1.12	1.00	915
SOCQ		0–60	42.65	11.31	0.82 **	915
GDS		0–14	2.90	3.07	−0.10 **	469
PFEIFFER		0–1	0.08	0.18	−0.22 **	473
BARTHEL		5–105	95.48	19.31	0.40 **	469
GHQ12		5–36	13.71	6.20	−0.36 **	420
UCLA		20–69	35.35	10.85	−0.55 **	418
MMSE		0.20–1	0.65	0.17	−0.020	200
GOLDBERG		1–2	1.72	0.39	0.05	200
EuroQol5D	**Total**	1–2.83	1.60	0.44	−0.45 **	200
	Health	1–10	6.46	2.02	0.47	199
EPQ-R	**E**	0–19	9.72	4.75	0.58 **	409
	N	0–23	11.08	5.23	−0.14 **	408
	p	0–16	4.83	2.90	−0.04	408

** The correlation is significant at the 0.01 level (bilateral).

As Table 4 shows, the correlation between both sociotype questionnaires was very high. In addition, the abbreviated scale had a high positive correlation with the extraversion subscale of the Eysenck EPQ-R, indicating that the higher the score in SOCG-4, the greater the extraversion; the relationship with the revised UCLA loneliness scale was also relevant, showing a strong negative correlation. The whole correlations of Table 4 will be discussed later.

Additionally, in Study 3, about the elderly population living at home versus those institutionalized in residences, we took the values corresponding to the 4 items of the abbreviated questionnaire and dichotomized them, obtaining the results below (Table 5):

Table 5. Comparative SOG-4 in the Study 3 population (residence vs. home dwelling).

	Provenance	*N*	Mn	Sd	Standard Error
SOCG-4	Home dwelling	100	3.00	1.19	0.12
	Residence	100	1.99	1.37	0.14

Student's *t*-test = 5.56; $p < 0.001$; effect size: Cohen's d = 0.787; and range of values 0–4.

4. Discussion

We obtained the new version of the target questionnaire of 4 items with 2 response options, by extracting 4 among the 12 original items and dichotomizing the original 6 response options from the SOCQ. This abbreviated SOCG-4 questionnaire for the geriatric population was put to trial in a pilot study, shown below in Appendix A.

In the sociodemographic data of our studies, men appear as underrepresented in the sample composition, as is common in many studies of the elderly population [32]. Most participants were widows/widowers (44.7%), and 38.6% lived alone—these two conditions become potential factors of perceived loneliness, specially the former. This was observed in the sociotype study for social relationships in the geriatric population [29], in which the widow/widower condition approximately cuts in half the average conversation time of the surviving partner. We can consider that living alone and being a widow represent two widespread sociodemographic features of the elderly population that, together with the low-income level (35.8% of our sample was below the average pension) and poor health, may have an important effect on social relationships. The geriatric facultative should carefully ponder these conditions as genuine risk factors of perceived loneliness, in addition to specialized questionnaires, such as the abbreviated ones observed herein.

The results presented in Tables 2 and 3 are evidence that the discriminant quality of the 12-item SOCQ in the general population may be mirrored in the elderly population as well. However, slight differences appear between both tables, which would be more significant if those differences could be extracted by specifically comparing the disjointed age groups (the general population of Study 1 contained close to 20% of the elderly population). The weights obtained for the different questions in the sociotype's three dimensions, show clear differences that precisely guided the selection procedure in the three phases already mentioned.

Regarding the four questions finally selected, they may be interpreted attending to their sociological scope. Question 1, "I talk and relate with my family", belongs to the emotional loneliness component, related to the presence or absence of intimate relationships with a partner or a best friend [30]. We may consider that it partially overlaps with Question 2, "I talk and relate with my friends", when strong bonds of friendship exist. However, in general, the emotional attachment to friends is lower than it is to partners—see the dramatic change in the talking time of the widows/widowers already mentioned [29]. The social loneliness aspect would then be covered by Question 2, and also by Question 3 (in reverse): "It is difficult for me to make conversation with people I do not know." This point on weak or nil relatedness is important. For the capability to uphold the mental effort that conversation with unknown parties usually implies (and the extra reward often obtained), seems to be a significant indicator of mental health and the personal potential for active aging and longevity. Remarkably, this was found by Julianne Holt-Lundstat in a series of relevant population studies [3,4]. Rather surprisingly (not so from the sociotype point of view), maintaining social relationships with unrelated individuals lists as the first factor for predicting longevity, immediately followed by close family relationships. All the usual predictors and conventional advice, such as exercising, quit smoking and drinking, diet, flu vaccination, and going to the doctor, appear later in the list. Clearly, complete social engagement, a rich sociotype, is a prerequisite for physical and mental health, especially for elderly persons. Therefore, it is important to rely on a simplified but valid instrument to assess this.

Concerning Question 4, "I have fun and laugh with my friends", the appearance of laughter among the selected questions may look odd. In fact, none of the existing questionnaires on the related topics (e.g., UCLA loneliness scale, MSPSS, SNI, Duke, SELSA, MOS, SSB, Zimet, and de Jong) include laughter. Is it an anomalous presence? Far from that, the research by some of the present authors [33–35] and by many others [36–39] indicate that laughter is closely related to the development of sociality and particularly to the formation and maintenance of social bonds. Furthermore, according to the population studies of Hassan and Hassan [38], the frequency of laughter is an excellent predictor of good physical health and mental health, as can be confirmed in medical stories themselves. In this research, once the three questions with the highest scores were selected, this fourth question centered on laughter obtained the best scores in its correlation with the other health questionnaires.

Beyond the excellent correlation with the SOCQ (0.82), in Table 4, the abbreviated dichotomous scale of SOCG-4 correlates very well with UCLA's detection of feelings of loneliness and social exclusion (−0.55) and with the GDS metrics of depression (−0.105), as well as with the extraversion subscale of Eysenck (0.58). It means that the abbreviated questionnaire gauges the feelings of loneliness quite soundly and captures very well the personal differences in extroversion. Concerning the acceptable value of the correlation with GHQ12 (−0.36), it indicates that poor health and psychological distress have an impact on social engagement, and vice versa. This result was not unexpected, for loneliness has already been proved to be a most important risk factor for both the physical health and mental health of the elderly [2,22,40]. The relationship with the scale of physical activities of daily life (Barthel scale) was also positive (0.30), with a better performance for people with high scores in autonomy. Finally, another remarkable correlation appears with the EuroQol-5D scores, also in the vicinity of 0.5 (−0.45 for distress in the quality of life and

0.47 for the global health index, respectively). All these significant correlations indicate the potential usefulness and validity of SOCG-4 as an auxiliary tool concerning the screening of the old population in geriatric clinics and primary care, with the obvious advantages of its very easy use and friendly questioning, always maintaining the necessary cautions when relying on a very simple, dichotomous scale. As an indicative example of its use, in Appendix A below, we describe the first fieldwork experience using this questionnaire in three different contexts: in phone attention, in home visiting, and in the clinic.

About the results in Table 5, obtaining SOCG-4 values for the elderly population living at home vs. those institutionalized in residences, the differences are very significant, indicating that the questionnaire is valid to discriminate between the two groups (Student's *t*-test and Cohen's d clearly indicate this). The standard error corresponds to the average misclassification error when only looking at the values of the SOCG-4 questionnaire. Moreover, in the population of Study 3, the correlation between SOCG-4 and SOCQ reaches 0.89, higher than the general 0.82 that appears in Table 4 for the combined population of all the studies.

Finally, regarding the relationship between the abbreviated sociotype questionnaire and other relatively similar tests, such as perceived social support [41], social-emotional loneliness [30], or the abbreviated UCLA [42], despite circumstantial resemblances, there is an important difference with them. The latter tests have an implicit sense of dependence, vulnerability, counting on alien support for covering personal needs, or even directly assuming a depressed state, while the sociotype refers to unmediated relationships, spontaneous talking, and a sense of empowerment while the subject carries her/his relationships autonomously. The subjects responded far more easily and with a better mood to the SOCQ and SOCG-4 questions, in particular to the question on laughter. Additionally, this user-friendliness was authenticated by social workers already using institutionally the SOCQ questionnaire (in comparison with UCLA), and by the pilot fieldwork use of SOCG-4, described in Appendix A. The results obtained in the fieldwork and the comments by the interviewers made clear the simplicity, usefulness and discriminating capability of the new questionnaire.

Thus, by comparing the performance of the new questionnaire with the other abbreviated tests, we may speculate that, in general, the degree of usefulness of say "positive" versus "negative" constructs would depend on the level of autonomy of the subject. For instance, in the case of relatively young elderly individuals, the positive construct would fare better as a tool for distinction, while for the oldest segments (or "fourth age") there would be an added capability to distinguish by the negative constructs. In any case, the correlations obtained for SOCG-4 with the other positive/negative tests in Table 4 look really promising, since it is a simplified, dichotomous test of only four questions. It is relevant that in the preliminary use of the new questionnaire, it has shown sufficient sensibility to establish the distinctions between the different populations in the three contexts.

5. Conclusions

The most important protection for the elderly, counting with a modicum of social relationships, is often left out of scope in current medical care, highly segmented and focused on organ failures and age-related diseases [21]. By paying special attention to social isolation and perceived loneliness, the primary care doctor or the geriatric facultative will be, in fact, monitoring the state of the most important risk-factor for the patient's health.

Thereafter, prescribing "socialization pills", via the coordination with municipal or regional day centers for the isolated elders, may raise the quality of life and in some cases prevent anti-depressive medication and the harmful syndrome of polypharmacy in the elderly individuals [21]. This implies the development of integrative management plans for general health, and the necessary interrelationship between medical care and social care institutions. It is not only the sustainability of health care systems that is at stake, but the betterment of daily lives for a growing population of increasingly isolated elders.

In this context, what the sociotype initiative means, and particularly SOCG-4, is the promotion of an easy instrument to be kept handy on the desk of the primary care and the geriatric clinic—acting as a reminder of the main health risk-factor of the old person standing in front of the doctor. We may insist on the importance of this new abbreviated metric, given the increase in loneliness after the COVID-19 pandemic.

Some of the SOCG-4 features may deserve ample consideration, such as simplicity, conviviality, user-friendliness, and inclusion of laughter. From a practical point of view, we explain in Appendix A how this fared in a phone survey, in home visiting, and in the clinic arenas. However, irrespective of the concrete performances of SOCG-4 and its basic limitations, it should stimulate discussion to contribute to a change of mentality in today's care system, refocusing on the socialization needs not so well covered amidst the frequent administrative and technological shifts. Rapidly aging societies across the world demand new orientations, new intervention strategies, and new screening instruments for the gerosciences. The post-pandemic society desperately needs this type of approaches.

Author Contributions: Conceptualization, P.C.M., J.N., A.G.-L. and F.J.O.; methodology, J.N., P.C.M., A.G.-L., M.G.-L. and F.J.O.; software, M.C. and J.N.; formal analysis, M.C. and J.N.; investigation, F.J.O.; data curation, P.C.M., M.C. and J.N.; writing—original draft preparation, J.N., P.C.M. and A.G.-L.; project administration, A.G.-L. and M.G.-L.; funding acquisition, A.G.-L. and M.G.-L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Ministerio de Economia y Competitividad (Spain), grantnumber CSO2017-82110-R.

Institutional Review Board Statement: The methodology and questionnaires of these studies had been previously approved by the Ethical Committee of Aragón (CEICA), Spain (Act: CP13/2014). In the case of Study 3, the response time was in the order of fifty minutes, and its approval was granted by the Ethical Committee of the Hospital Complex of Navarra (Act: 18 January 2017).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: The cooperation of ATENZIA (a social care company in Zaragoza) is acknowledged regarding the use of SOCG-4 in the phone survey and home visiting. We also thank to Laura Montes for her cooperation in field work.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Appendix A. Use of SOCG-4

SOCG-4 was put to use in three different contexts: phone survey, home visiting, and psychogeriatric clinic. The two former cases were in charge of specialized employees of a social work company, while the latter case was in charge of a psycho-geriatric facultative, one of the authors of this paper, at San Jorge Hospital (Huesca, Spain). A total of 150 subjects were administered the abbreviated scale: 50 subjects in each modality. The social workers involved in the interviews were naïve regarding SOQG-4, and they did not receive any previous directions on how to administer the questionnaire and how the questions themselves should be asked or complemented. A casual mode, embedded within the usual procedure, was the only suggestion. Records were minimized or suppressed, and just the responses to the questions were consigned (no names, no indicators, no sociodemographic data, and no other tests). Apart from consigning the responses, only a qualitative impression was obtained. In the clinic case, age and sex were also annotated by the facultative.

Appendix A.1. Comments from Interviewers

Phone use. Questions were easy to formulate by the specialized personnel (obviously in all cases the questions had to be changed from first person to second person). Most parties responded positively to the first question about family, and also to Question 2 about friends, but not without some hesitation or shame when their response was, or should have been, negative. Being asked about friendly relationships without face-to-face report made this question more difficult. In general, the response to this question was closely related to Question 4 on laughter. Nevertheless, administering the test was useful to provide a rough picture of the social relationships of those interviewed and, frequently, it also facilitated engaging in further phone conversations by the interviewer.

Home use. In this case, the difference between "preventive" users of the home-visiting service and those involved in the "dependence" official program was evident. The latter suffered from pathologies or physical/mental handicaps that restricted their capability for autonomous deambulation and outdoor activities. The positive response to the relationship with family was stronger in these dependent parties, while they easily acknowledged their missing social relationships—but not in the case of the preventive users of the service. An interesting gender difference was noted by the interviewer in charge of the survey; regarding Question 3, difficulty with strangers, men were responding more directly and positively than women. At stake is whether it was more a matter of maintaining a conventional feminine self-image or a genuine behavioral preference.

Clinic use. It was a very positive experience from the beginning. Older people who came to a consultation for problems of depression, cognitive impairment, or anxiety were pleasantly surprised when asked about family and friends. However, particularly regarding Question 1, about family, some of them were trying to please the facultative. For example, those who lived in a residence often commented that they received frequent visits from relatives and this was not really the case—the residence companions told the facultative about that. Those patients did not like to cast a negative image on themselves regarding their family bonds. In the other questions, that kind of obstacle was not observed and there was less trouble to show the degrees of personal isolation. In Question 3, for instance (difficulty with strangers), they answered that this was indeed the case. Older people had more difficulty in "opening up" to unknown parties, perhaps due to more fear and insecurity. In Question 4 (laughter), having fun and laughing with friends was complicated for the adults with depression or dementia. This question often caused them certain sadness when they found out that this was not their case. The affective state also influenced the responses. The depressed and sad individuals tended to respond that they did not talk or relate to family and friends, while their companions often commented that it was more "a feeling" than a reality. Overall, the test worked very well and most patients had a positive sensation of increased attention and personal care by the facultative.

Appendix A.2. Quantitative Results

The results in Table 1, expressed in %, represent the positive responses to the different questions, with the proportion of "1s" obtained. In Question 3, the response is reversed (counting then the "0es"). Among the three contexts, the geriatric clinic use of the test appeared as the most sensitive, by far. With respect to Q1, however, the reluctance to accept poor family relationships was remarkable in all instances. Thus, a good suggestion could be the use of a complementary question just after the positive response (for instance, "Have you talked with them this week or last week?" or, "When was the last time you talked with them?").

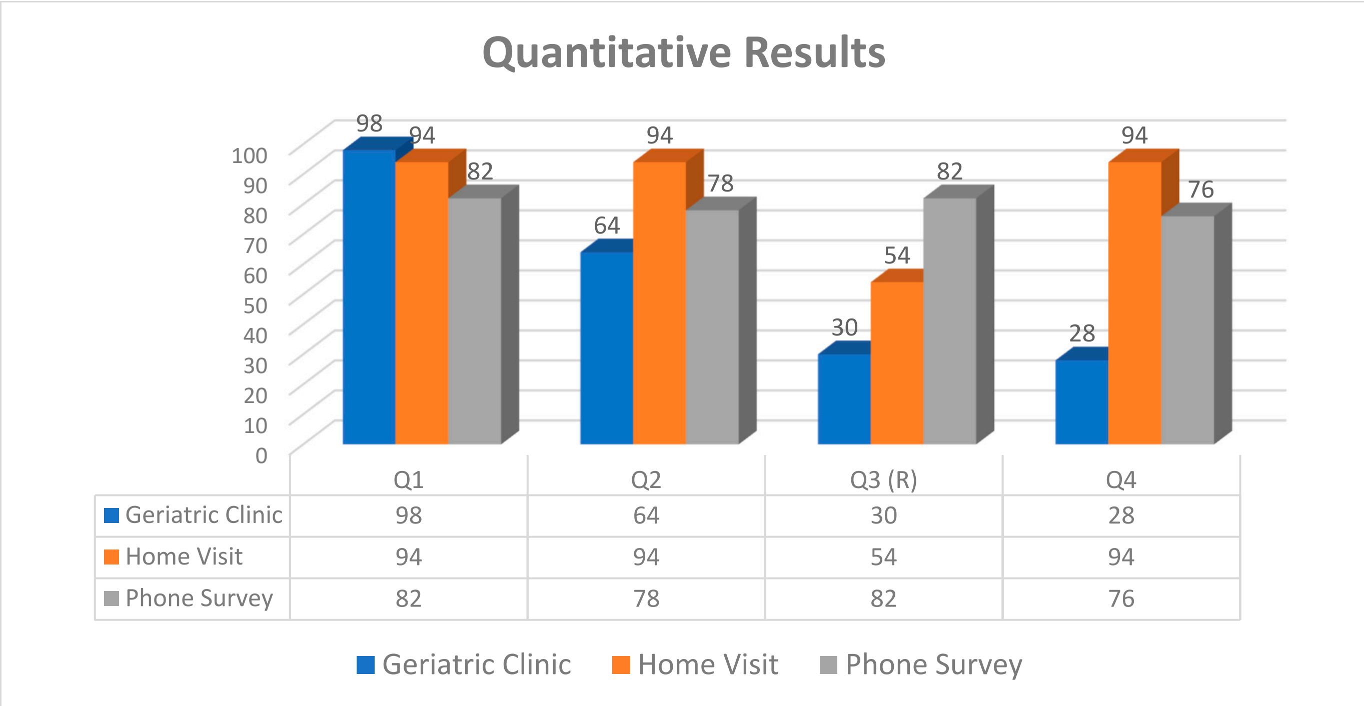

Figure A1. Quantitative results of SOCG-4 in three different contexts: geriatric clinic, home visit, and phone survey.

Regarding the discrimination of patients with a relative risk of isolation, there were noticeable differences between the outcomes of the different contexts. The clinic use was producing risk signals in 31 cases out of 50 (counting two or more "0es"), while in the other 2 contexts the warning signals were, respectively, 11 (phone) and 5 (home visiting). The disparaging results in Questions 3 and 4, in between the three contexts, might be partially explained by the relatively worsened conditions of geriatric patients and home visiting users. In any case, this was only a preliminary use of the test; rigorous fieldwork should be developed alongside future research by the authors.

References

1. Perissinotto, C.M.; Covinsky, K.E. Living Alone, Socially Isolated or Lonely—What are We Measuring? *J. Gen. Intern. Med.* **2014**, *29*, 1429–1431. [CrossRef] [PubMed]
2. Cacioppo, J.; Cacioppo, S. Loneliness Is a Modern Epidemic in Need of Treatment. 2014, Volume 19, p. 2016. Available online: https://www.newscientist.com/article/dn26739-loneliness-is-a-modern-epidemic-in-need-of-treatment/ (accessed on 20 January 2022).
3. Holt-Lunstad, J.; Robles, T.F.; Sbarra, A.D. Advancing social connection as a public health priority in the United States. *Am. Psychol.* **2017**, *72*, 517. [CrossRef] [PubMed]
4. Holt-Lunstad, J.; Smith, T.B.; Baker, M.; Harris, T.; Stephenson, D. Loneliness and Social Isolation as Risk Factors for Mortality: A Meta-Analytic Review. *Perspect. Psychol. Sci.* **2015**, *10*, 227–237. [CrossRef] [PubMed]
5. Aronson, N. *Elderhood: Redefining Aging, Transforming Medicine, Reimagining Life*; Bloomsbury Publishing: London, UK, 2019.
6. WHO. *Mental Health and Psychosocial Considerations during COVID-19 Outbreak*; World Health Organization: Geneva, Switzerland, 2020; pp. 1–6.
7. Groarke, J.M.; Berry, E.; Graham-Wisener, L.; McKenna-Plumley, P.E.; McGlinchey, E.; Armour, C. Loneliness in the UK during the COVID-19 pandemic: Cross-sectional results from the COVID-19 Psychological Wellbeing Study. *PLoS ONE* **2020**, *9*, e0239698. [CrossRef] [PubMed]
8. Granovetter, M. The impact of social structure on economic outcomes. In *The Sociology of Economic Life*, 3rd ed.; Routledge: Oxfordshire, UK, 2018. [CrossRef]
9. McPherson, M.; Smith-Lovin, L.; Brashears, M.E. Social isolation in America: Changes in core discussion networks over two decades. *Am. Sociol. Rev.* **2006**, *71*, 353–375. [CrossRef]
10. Larg, A.; Moss, J.R. Cost-of-illness studies: A guide to critical evaluation. *PharmacoEconomics* **2011**, *29*, 653–671. [CrossRef]

11. Mihalopoulos, C.; Le Khanh-Dao, L.; Chatterton, M.L.; Bucholc, J.; Holt-Lunstad, J.; Lim, M.H.; Engel, L. The economic costs of loneliness: A review of cost-of-illness and economic evaluation studies. *Soc. Psychiatry Psychiatr. Epidemiol.* **2020**, *55*, 823–836. [CrossRef]
12. Putnam, R. *Bowling Alone*; Simon & Schuster: New York, NY, USA, 2000.
13. Shankar, A.; McMunn, A.; Banks, J.; Steptoe, A. Loneliness, Social Isolation, and Behavioral and Biological Health Indicators in Older Adults. *Health Psychol.* **2011**, *30*, 377. [CrossRef]
14. Shor, E.; Roelfs, D.J.; Yogev, T. The strength of family ties: A meta-analysis and meta-regression of self-reported social support and mortality. *Soc. Netw.* **2013**, *35*, 626–638. [CrossRef]
15. Qualter, P.; Vanhalst, J.; Harris, R.; van Roekel, E.; van Roekel, E.; Lodder, G.; Bangee, M.; Maes, M.; Verhagen, M. Loneliness Across the Life Span. *Perspect. Psychol. Sci.* **2015**, *10*, 250–264. [CrossRef]
16. Booth, R. Loneliness as a component of psychiatric disorders. *Medscape Gen. Med.* **2000**, *2*, 1–7.
17. Grant, N.; Hamer, M.; Steptoe, A. Social isolation and stress-related cardiovascular, lipid, and cortisol responses. *Ann. Behav. Med.* **2009**, *37*, 29–37. [CrossRef] [PubMed]
18. Cacioppo, J.T.; Cacioppo, S.; Capitanio, J.P.; Cole, S.W. The neuroendocrinology of social isolation. *Annu. Rev. Psychol.* **2015**, *66*, 733–767. [CrossRef] [PubMed]
19. Ruan, H.; Wu, C.F. Social interaction-mediated lifespan extension of Drosophila Cu/Zn superoxide dismutase mutants. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 7506–7510. [CrossRef] [PubMed]
20. Doane, L.D.; Adam, E.K. Loneliness and cortisol: Momentary, day-to-day, and trait associations. *Psychoneuroendocrinology* **2010**, *35*, 430–441. [CrossRef]
21. Campisi, J.; Kapahi, P.; Lithgow, G.J.; Melov, S.; Newman, J.C.; Verdin, E. From discoveries in ageing research to therapeutics for healthy ageing. *Nature* **2019**, *571*, 183–192. [CrossRef]
22. Berkman, L.F. Social epidemiology: Social determinants of health in the United States: Are we losing ground? *Annu. Rev. Public Health* **2009**, *30*, 27–41. [CrossRef]
23. Perissinotto, C.M.; Cenzer, I.S.; Covinsky, K.E. Loneliness in older persons: A predictor of functional decline and death. *Arch. Intern. Med.* **2012**, *172*, 1078–1084. [CrossRef]
24. Coyle, C.E.; Dugan, E. Social isolation, loneliness and health among older adults. *J. Aging Health* **2012**, *24*, 1346–1363. [CrossRef]
25. Luo, Y.; Waite, L.J. Loneliness and mortality among older adults in China. *J. Gerontol. Ser. B Psychol. Sci. Soc. Sci.* **2014**, *69*, 633–645. [CrossRef]
26. Rowe, J.W.; Fulmer, T.; Fried, L. Preparing for Better Health and Health Care for an Aging Population: A Vital Direction for Health and Health Care. *NAM Perspect.* **2016**, *6*, 1643–1644. [CrossRef]
27. Marijuán, P.C.; Montero-Marín, J.; Navarro, J.; García-Campayo, J.; del Moral, R. The 'sociotype' construct: Gauging the structure and dynamics of human sociality. *PLoS ONE* **2017**, *12*, e0189568. [CrossRef] [PubMed]
28. Dunbar, R. *The Human Story*; Faber and Faber: London, UK, 2004.
29. Marijuán, P.C.; del Moral, R.; Ji, S.; Lacruz, M.G.; Gómez-Quintero, J.D.; Navarro, J. Fundamental, Quantitative Traits of the 'Sociotype'. *BioSystems* **2019**, *180*, 79–87. [CrossRef] [PubMed]
30. Gierveld, J.D.J.; van Tilburg, T. A 6-item scale for overall, emotional, and social loneliness: Confirmatory tests on survey data. *Res. Aging* **2006**, *28*, 582–598. [CrossRef]
31. Cornwell, E.Y.; Waite, L.J. Measuring social isolation among older adults using multiple indicators from the nshap study. *J. Gerontol. Ser. B Psychol. Sci. Soc. Sci.* **2009**, *64*, i38–i46. [CrossRef]
32. Domènech-Abella, J.; Lara, E.; Rubio-Valera, M.; Olaya, B.; Moneta, M.V.; Rico-Uribe, L.A.; Ayuso-Mateos, J.L.; Mundó, J.; Haro, J.M. Loneliness and depression in the elderly: The role of social network. *Soc. Psychiatry Psychiatr. Epidemiol.* **2017**, *52*, 381–390. [CrossRef]
33. Marijuan, P.C.; Navarro, J. The Bonds of Laughter: A Multidisciplinary Inquiry into the Information Processes of Human Laughter. *arXiv* **2010**, arXiv:1010.5602.
34. Navarro, J.; Rosell, M.F.; Castellanos, A.; del Moral, R.; Lahoz-Beltra, R.; Marijuán, P.C. Plausibility of a Neural Network Classifier-Based Neuroprosthesis for Depression Detection via Laughter Records. *Front. Neurosci.* **2019**, *13*, 267. [CrossRef]
35. Navarro, J.; del Moral, R.; Alonso, M.F.; Loste, P.; Garcia-Campayo, J.; Lahoz-Beltra, R.; Marijuán, P.C. Validation of laughter for diagnosis and evaluation of depression. *J. Affect. Disord.* **2014**, *160*, 43–49. [CrossRef]
36. Martin, R.A. Is laughter the best medicine? Humor, laughter, and physical health. *Curr. Dir. Psychol. Sci.* **2002**, *11*, 216–220. [CrossRef]
37. Uekermann, J.; Channon, S.; Lehmkämper, C.; Abdel-Hamid, M.; Vollmoeller, W.; Daum, I. Executive function, mentalizing and humor in major depression. *J. Int. Neuropsychol. Soc.* **2008**, *14*, 55–62. [CrossRef]
38. Hasan, H.; Hasan, T.F. Laugh yourself into a healthier person: A cross cultural analysis of the effects of varying levels of laughter on health. *Int. J. Med. Sci.* **2009**, *6*, 200. [CrossRef] [PubMed]
39. Takeda, M.; Hashimoto, R.; Kudo, T.; Okochi, M.; Tagami, S.; Morihara, T.; Sadick, G.; Tanaka, T. Laughter and humor as complementary and alternative medicines for dementia patients. *BMC Complementary Altern. Med.* **2010**, *10*, 1–7. [CrossRef] [PubMed]
40. Klinenberg, E. Social Isolation, Loneliness, and Living Alone: Identifying the Risks for Public Health. *Am. J. Public Health* **2016**, *106*, 786–787. [CrossRef] [PubMed]

41. Zimet, G.D.; Dahlem, N.W.; Zimet, S.G.; Farley, G.K. The Multidimensional Scale of Perceived Social Support. *J. Personal. Assess.* **1988**, *52*, 30–41. [CrossRef]
42. Hughes, M.E.; Waite, L.J.; Hawkley, L.C.; Cacioppo, J.T. A short scale for measuring loneliness in large surveys: Results from two population-based studies. *Res. Aging* **2004**, *26*, 655–672. [CrossRef]

BioSystems 233 (2023) 105039

Contents lists available at ScienceDirect

BioSystems

journal homepage: www.elsevier.com/locate/biosystems

Natural intelligence and the 'economy' of social emotions: A connection with AI sentiment analysis

Jorge Navarro [a,*], Pedro C. Marijuán [b]

[a] Grupo de Decisión Multicriterio Zaragoza (GDMZ), Faculty of Economics, University of Zaragoza, 50006, Zaragoza, Spain
[b] Independent Scholar Affiliated to Bioinformation Group, Aragon Health Science Research Institute (IIS Aragon), Zaragoza, Spain

ARTICLE INFO

Keywords:
Natural intelligence
Life cycle
Social emotions
Circumplex model
Sentiment analysis

ABSTRACT

By approaching the concept of Natural Intelligence a new path may be open in a variety of theoretical and applied problems on social emotions. There is no doubt that intelligence emerges as a biological/informational phenomenon, although paradoxically a consistent elaboration of that concept has been missing. Regarding emotions, they have been keeping an unclear status, being often restricted to the anthropological or to ethological approaches closer to the behaviorist paradigm. Herein we propose a different track, centered in the life cycle advancement. The life cycle in its integrity becomes the nucleus of natural intelligence's informational processes, including the consistent expression of emotions along the maximization of fitness occasions. In human societies, the overall 'economy' of social emotions is manifest, showing up in the conspicuous interplay between bonding processes and different classes of social emotions. The essential link between natural intelligence, emotions, and the life cycle of individuals may harmonize with current progresses – and blind spots – of artificial intelligence fields such as 'sentiment analysis.'

1. Introduction: Approaching natural intelligence

One way to initially approach natural intelligence in its relationship with social emotions could be placing a contraposition between the two kinds of intelligence: natural and artificial. In this contraposition, emotions appear as one of the fundamental components of natural intelligence, but at the same time they constitute one of the most cherished targets of socially applied artificial intelligence—the core of the "attention economy (Lanham, 2006). In our approach to social emotions from the point of view of the former, we will largely benefit of recent biological achievements, trying to translate into the human societies a consistent conceptualization based on the cellular life-cycle's interception of information flows. We will show that our informational nature of composite "cellular individuals" involves a parsimonious deployment of emotions in the individual achievement of fitness within the social milieu. The challenge is how to establish a compact narrative that might lead, so to speak, from living cell-cycles to behaving animal brains and human individuals within social structures. That's precisely the central goal of this work.

Let us clarify first what we mean by the rather infrequent term of 'natural intelligence.' It may be ascertained, or better intuited, via different multidisciplinary syntheses among a plurality of new fields that have progressively emerged along the biomolecular and computer revolution of recent decades: evolutionary epistemology, autogenesis, autopoiesis, bioinformation, biological cognition, biocybernetics, biosemiosis, natural computation, bioinformatics, biocomputing, bioengineering, synthetic biology, systems biology, and so on. Different synthetic views about some of these fields may be found in (Armitage et al., 2005; Corning, 2020; Marijuán, 2002; Marijuán and Navarro, 2022; Perez Velazquez, 2009; Shklovskiy-Kordi and Igamberdiev, 2022; Slijepcevic, 2018; Timsit and Grégoire, 2021; van Duijn, 2017). Let us note that the term intelligence frequently appears in these works, referred to information processing and often extended to cells, multicellulars, plants, nervous systems, swarms, and animal societies (Calvo, 2016; Gershenson, 2021; Solé et al., 2019; Trewavas, 2017). But the conceptual panorama is far from coherent. Indeed, intelligence participates of the conceptual difficulties of a series of deeply related concepts such as information, meaning, value, and knowledge. In fact, whatever the field, the concept becomes unassailable, and counts like its germane information with multiple pragmatic definitions – too many of them! – related to the traditional fields in which it has been applied: psychology, social science, cognitive science, computer science, etc.

* Corresponding author.
E-mail address: jnavarrol@unizar.es (J. Navarro).

https://doi.org/10.1016/j.biosystems.2023.105039
Received 8 August 2023; Received in revised form 21 September 2023; Accepted 21 September 2023
Available online 22 September 2023

Pragmatically, intelligence means the ability to flexibly use different tools in the pursuit of some adaptive goals in a changing environment. In the field of AI, as indicated by its pioneers (McCarthy et al., 1955), the interest on intelligence was just pragmatic. They were interested in automatic computers, language programming, neural networks, complexity measurement, abstraction processing, randomness, and creativity. Very soon, however, the success of this venture lead to the "cognitive revolution," which proposed a hexagon of disciplines focused on human cognition and artificial intelligence: linguistics, neuroscience, artificial intelligence, philosophy, anthropology, and psychology. Any biological connections were disregarded; it was a separate kind of intelligence, only anthropocentric and computational.

Nevertheless, as one of the present authors argued long ago, in the early 1980s, the exclusively formal, computer-based schemes of expert and logical systems, perceptrons, neural networks, parallel processing, etc., were far from sufficient. Previously, a reflection on the general phenomenon of intelligence in nature was needed, also taking distances from too anthropocentric approaches in psychology or from behaviorism-oriented ethology. These ideas were developed as a PhD thesis: *"Natural Intelligence: The Evolution of Biological Information Processing"* (Marijuán, 1989). To make a long story short, the scheme of this work was oriented bottom-up, from enzymes as basic processors (molecular automata), coupled with memory banks (nucleic acids), to the informational-intellective scheme of living cells; going then to multicellularity, to the emergence of nervous systems, and to fundamentals of intelligence based on neural processing; ending with an approach to crucial aspects of social intelligence. The commonality of features among highly different, but hierarchically interrelated, forms of intelligence was dubbed as 'natural intelligence' (Marijuán, 1989). This approach pointed to some fundamentals of cellular, neural, and social collective intelligence that, overall, would continue to be applicable and of some theoretical interest, also facilitating the connection with social emotions.

Properly establishing the core of natural intelligence is too complex a multidisciplinary task. Just trying to model or simulate in AI grounds some of its multifarious capabilities does not seem to conduce to the core problems (Bryson, 2015; Gershenson, 2021). Herein, in order to connect with social emotions, we will make a detour around a series of ideas that establish a workable scheme of natural intelligence. It will start, in Section 2, with the cellular world, taking prokaryotic intelligence as the 'fundamental unit (Armitage et al., 2005; Marijuán et al., 2010). An essential point will be advanced: the primacy of the cellular *life cycle* (life story, life course) as the main subject of evolutionary information processing and adaptation (Minelli, 2015). Analyzing the behavioral developments of the cell and its interception of information flows, we will find a curious simile of, say, 'molecular emotions' at the level of gene expression—around the functional hierarchy of *sigma factors*. Parallel considerations around the life cycle of multicellular organisms will guide in our analysis of the adaptive information-processing role of nervous systems. We will register the appearance of true emotions along the evolution of nervous systems, and we will also consider the optimality or 'economy' principles of neuronal information processing (Friston, 2012; Tozzi et al., 2017)—not to forget that it was Santiago Ramón y Cajal (1899) who first proposed 'principles of economy' in nervous system anatomy and physiology.

Previous to the discussion in depth of the human emotional endowment, in Section 3 we will focus on the stringent social adaptation of our species – the 'niche' structure of our own sociality – what can be called the human "sociotype", and the influence that the different relational domains therein would have on our own arrangement of social emotions (Marijuán et al., 2017, 2019). Then, in Section 4, we will approach the Cartesian diagram of emotions, the *circumplex model*, which represents valence and activation (POSNER et al., 2005; Russell, 1980; Russell and Barrett, 1999); other classifications of emotions will be revised but, rather than emphasizing the making of longer and longer all-encompassing emotion lists, we will propose the systematic analysis of the patterns and coalition patterns that trigger off our social emotions. It is an idea that can also be related to the emergence of human "ultrasociality" (Turchin, 2016). The basic or primary emotions involved in processes of *strong bonding*, as well as the secondary emotions involved in intergroup cooperation and *weak bonding*, would be systematically co-opted for the emergence and structural consolidation of ultrasociality. There seems to be a global interrelationship of new bonding processes and appearance of new social emotions, crystallized in very different cultures, social structures, and institutions.

Summing up at the Conclusions Section, and closing the conceptual circle, the whole case of social emotions will be reconsidered in the light of the dichotomy between natural intelligence and artificial intelligence—and their contemporary collision in 'sentiment analysis'.

2. The fundamental unit of natural intelligence

It is important that we clarify the origins of biological/natural intelligence and its cellular 'fundamental unit', which is necessary to set up a consistent rationale along the evolutionary process. In other words, there appears a 'prehistory' of intelligence – and of emotions –which undoubtedly is cellular. Its evolutionary development provides a deep sense and cogency to the social emotions approach.

2.1. The cellular life cycle

Cellular systems (even the simplest ones, prokaryotes) purport an amazing information design. The living cell is a system that constructs itself from environmental material according to an internal blueprint that is separate from the constructive system itself (following von Neumann's self-reproducing automata). All external substances used for self-construction are systematically detected and identified by a dedicated apparatus, the signaling system; then they are selectively imported into the cytoplasm in order to extract their free energy along the ensuing metabolic pathways. So, metaphorically speaking, 'reading' the environment affordances becomes prior to 'eating' them. Or more conventionally, the high-energy, highly valuable energy flows apportioning the materials needed for self-production will be anticipated, detected, and captured by means of the faster and cheaper communication flows tended with the surrounding environment via the cellular signaling system. Recent discoveries in prokaryotic cellular signaling systems have evidenced new important details of this sophisticate relational phenomenon (Galperin, 2005; Grigoroudis et al., 2007; Ulrich et al., 2005). What has been called the "1-2-3 Component Systems" scheme (as coined by Jorge Navarro in (Marijuán et al., 2010) has opened new views on a variety of themes related to the cellular integration of signaling with metabolism and gene transcription—the core of biological intelligence.

An important aspect, looking at the way the bacterium E. coli – the most traditional biomolecular model system – organizes its gene transcription processes, is that a few general states strongly orientate the whole activities: favorable growth conditions, thermal and osmotic stresses, starvation, lack of iron, etc. A few "sigma" factors take care of responding to such specific conditions. In E. coli there are 7 different "sigmas", which are respectively known by their weigh in kDalton. They link RNA polymerases to transcription factors and gene promoters. One of them, Sigma 70, which is constitutively expressed, correlates with favorable growth conditions, and promotes generic translation of around 40% of the genome; the other sigma factors directly promote the expression of around one hundred genes or less. They may cover desiccation, starvation, iron presence, sporulation, SOS system, etc. In all cases, these sigma factors are carefully controlled: anti-sigma proteins and anti-*anti*-sigma proteins take care of detecting specific state variables that determine their gene expression (Gama-Castro et al., 2016; Karp et al., 2007; Salgado et al., 2013).

Let us emphasize that the sigma factors' role in the bacterial life cycle parallels the role of emotions in central nervous systems—propitiating a complete reorientation of cellular behavior by switching towards another aggregate of molecular pathways which are more favorable for the advancement of the life cycle.

If we focus on this continuous relationship with the environment, the adaptive responses that the cell synthesizes may now be contemplated under the prism of what the signal 'invisibly' conveys—its meaning. Meaning emerges from the cell's capability to self-modify its structures in response to external signals, and to inner changes as well. We may state that the meaning of a signal is what is fabricated ad hoc by the receiver cell, essentially via its signaling system and the coupled protein synthesis. This is a universal trait maintained in all the kingdoms of life.

Thus, we advocate the centrality of the cellular signaling system as the genuine source of biological semiosis along the evolutionary process (Marijuán et al., 2018). It is the capability of responding to external signals, changing the own structure via the relatively 'blind' self-production mechanisms, what supports the evolution of more advanced interactions and communicative exchanges—including 'molecular languages and all sort of intercellular codes. A cluster of concepts tightly associated with information and meaning, such as memory, value, and knowledge, all of them integrated within the whole of biological intelligence, may be suitably approached along this way of thinking (Marijuán and Navarro, 2022; Navarro and Marijuán, 2022).

Let us emphasize that while we have described the foundations of cellular intelligence, we have also found a 'protoemotional' system capable of dramatically redirecting the ongoing cellular life cycle towards its most adaptive course.

2.2. *Complex multicellular organisms: emergence of new forms of biological intelligence*

The different types of intelligence that have evolved in eukaryotic cells, multicellular organisms, fungi, plants, animals, etc. would not depart from the basic phenomenology of prokaryotes: their intellective mechanisms are always in the service of advancing the life cycle. New powerful developments such as symbiosis, signal expansion, cell cycle modularity, differentiation, epigenesis, and the ontogenetic development of metazoans did conduce to evolutionary scenarios of uncanny complexity. Eukaryotic cells developed new tools to control the possible paths of the now far more complex life cycle. It becomes in Borges' terms, "the garden of the forking paths" (Borges, 1941). Now, the functional equivalents of sigma factors are the cellular "checkpoints", where fundamental internal and external pieces of information converge to take the great decisions of a life cycle, which has now become extremely more complex: growth, maintenance, arrest, reproduction, differentiation, specialized function, apoptosis … All these complex checkpoints represent genuine "pattern recognition" devices the mission of which is to maintain or to transform the trajectory of the eukaryotic cell cycle along its most appropriate path. See Fig. 1.

As supracellular organisms come forward, endowed with multiple tissues playing a collective problem-solving game, they are still based on the signaling system capabilities of individual cells and their malleable life cycles. The nervous system is a case in point. It appears as a special electro-molecular tissue capable of orchestrating a new way of 'topodynamic' information processing, providing the body with instant fitness assessment based on increasing varieties of information acquisition (Tozzi et al., 2017). Electricity, ad hoc molecules, and topological mappings become the basic tools of this advanced form of biological

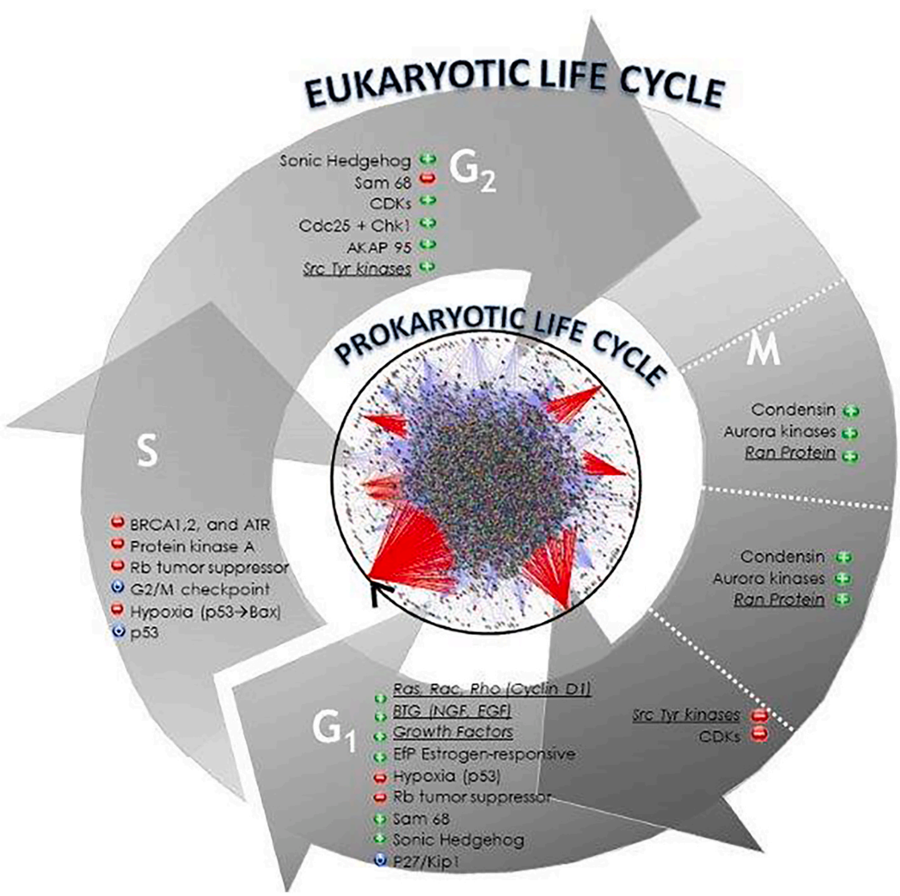

Fig. 1. Eukaryotic and prokaryotic life cycles: A comparative is drawn between prokaryotic sigma factors (fanning red lines in the center of the figure) and eukaryotic checkpoints (represented as dotted signs along the different phases of the life cycle).

intelligence. The evolution of neuronal intelligence has kept pace with the progressive complication and refinement of these nested information flows across the hierarchy levels of organization (Wurtz, 2021). From diffuse neural networks, to ganglia, to cords, to cerebroids, and to the central nervous systems of vertebrates, what we see is an informational crescendo culminating in advanced mammals and anthropoids where individuals may be organized not only into ecosystem-based dispersed networks, but also into coherently bonded societies.

Adaptability defines and sets the scope of natural intelligence. Successful adaptive action, which leads an organism and its genetic partners in the pursuit of long-term fitness, becomes the litmus test for any intelligent behavior. No matter how complex neural processes operate in complex organisms, they must always serve to drive the life cycle ahead, to maximize the organism's fitness. However, what could be the inner equivalent of fitness maximization within this new neuronal realm? A general principle of free energy minimization among the coupled neuronal discharges has been postulated, irrespective of the multiplication of mappings, localizations, and neurotransmitters, in order to organize adaptive behavior (Friston, 2012). An early approach to knowledge automation within central nervous systems, based on entropy minimization, was already proposed by Kenneth Paul Collins (Collins, 1991; Collins and Marijuán, 1997).

Once these complex nervous systems are at work with their sophisticate and ultrafast information processing, the open-ended detection of events coupled with the internal states and the action possibilities all must be integrated to appropriately pursue the fitness occasions. Reflexes, proto-emotions, and emotions are an essential part of the whole integrative process. The implementation mechanisms developed along the evolution of nervous systems basically go from relatively simple "fixed detection-action patterns" to more articulate "flexible" ones, and finally to elaborate "perception-action-reference superstructures" (Csany, 1988). It could remind the complexity jump we have found at the cellular level from sigma factors to cellular checkpoints. As a result of the different pattern recognition devices, either fixed or flexible or superstructural, a battery of neurotransmitters, neuromodulators, and neurohormones with different time scales and providing a variable interconnection among the different cortical areas and basal structures would have the capability to suspend the ongoing minimization processes and focus on the pursuit of the detected new fitness occasions.

We call 'emotions' to these processing superimpositions that interrupt the secondary and enforce the fundamental, always to establish the non-negotiable primacy of the life cycle. Then, the parallel with prokaryotic sigma factors and eukaryotic checkpoints seems appropriate: the ongoing life cycle recalibrates its inner processes and explores a new, more favorable trajectory for the sake of its own advancement. That's the overall mission of the whole mechanisms of natural intelligence: the maximization of fitness occasions, inclusively understood.

Researchers who have worked on the interplay between rationality and emotions in the human case would not be too far from this idea on the emotions' role of extending the information processing of life cycles towards the most adaptive directions (Arbib and Fellous, 2004; Damasio, 1994; Kahneman, 2011; Panksepp, 2015). Cogent rationality requires emotional support, guidance, and regulation (Koole, 2009). So to speak, being a sort of closed 'formal' processing system, the conscious reflective mind (Igamberdiev, 2023) needs to receive its processing goals and evaluations from afar.

The pertinence of this natural intelligence approach to emotions, now assuming all the complexity of human brain evolution and the interrelated social scenarios, will be addressed in the next Section.

3. The social environment of human emotions

The linguistic ability of human species has led our societies down a whole new path. The role of emotions, or better of the newfangled "social emotions", in the context of extended sociality, or even "ultrasociality" (Turchin, 2016), is now mediated by linguistic exchanges in larger and larger groups. Some prior conceptualizations in evolutionary anthropology and evolutionary psychology would be in order.

- The "*social brain*". We have evolved our big brains to cooperate, compete, and communicate in large, but close-knit, "natural groups". There seems to be an average of social networking, with rather ample upper and lower limits, concerning the number and types of bonding relationships that an individual can maintain meaningfully. The empirical finding of networking regularities such as the famous "Dunbar's number" (150–200 individual acquaintances) would make evolutionary and anthropological sense (Dunbar, 2004, 2007). These findings, integrated within the "social brain hypothesis", which was originally known as the Machiavellian intelligence hypothesis (Whiten and Byrne, 1988, Whiten and Byrne, 1989), show an ample clutch on the roots of human sociality and the origins of language. Essentially, this social brain hypothesis has posited that, in primate societies, selection has favored larger brains and more complex cognitive capabilities as a mean to directly cope with the challenges of social life (Allman, 1999; Silk, 2007). Subsequently, the overall cortical conformation and capacity of our species, vastly enlarged regarding other Anthropoidea, would sustain the high number of bonds that, comparatively, human individuals can maintain meaningfully within their oversized groups.
- The "*Sociotype*". As a result of the functioning of our social brain, despite all the existing cultural diversity, there is a great similarity of social relationships that has evolutionary roots ('descended from our genes'). The sociotype would represent our adaptive sociality, a relational whole consisting of a few characteristic sectors in nowadays societies: close family and kin circles, friends, work colleagues, and general acquaintances (Marijuán et al., 2017, 2019; Navarro et al., 2022). In the same way that there is scientific consensus on the validity of the *genotype* and *phenotype* constructs for the human species, notwithstanding their respective degrees of variability, a *sociotype* metrics could also be developed applying to the relative constancy of the social environment to which the individuals of our species would be evolutionarily adapted. The empirical quest by the authors has shown relational results that are relatively similar, and not far away from Dunbar's number in most cases, but with relevant differences (Ji, 2017; Marijuán et al., 2019). There is also a more holistic interpretation of the sociotype covering the influence of cultural backgrounds with a special focus in the individual's mental and physical health (Berry and De Geest, 2012; Berry, 2011). We need instinctive responses to achieve individual adaptation to the different kinds of groups, to achieve and maintain our own sociotype—and to achieve some level of collective "social intelligence" both intragroup and intergroup (Henrich, 2016). And we also need the reflective capabilities of the conscious mind, as outlined by V.A. Lefebvre regarding the use of language for social communication and information interaction (Igamberdiev, 2023).
- *Social emotions.* In a first approach to social emotions, we may consider that the 'big six' emotions traditionally discussed by theorists (Barrett, 2006; Ekman, 1992, 1999) are the most important in terms of their facial and bodily expressions (sadness, happiness, fear, anger, surprise, and disgust), but this does not mean they are most common or significant in our daily lives within the different social environments or in online communication. Rather we may find more often a series of dual sentiments and emotional reactions linked to group situations such as: exclusion vs. inclusion, sympathy vs. antipathy, admiration vs. envy, reward vs. punishment, irritability vs. calmness, excitement vs. composure, etc. We will consider some of these conditions in Section 4, within the new framework we are exploring for social emotions. Further, current AI works on sentiment analysis via lexicons are trying to analyze some of these dual reactions, regularly processing them via different procedures (Turón et al., 2023). The interrelationship of social networks texts with

emotions is also an essential matter for commercial platforms (Zuboff, 2019).

Too many open questions remain. Among them: How social emotions relate with the making and breaking of social bonds (and, particularly, with what kinds of social bonds)? How social emotions could be analyzed and classified in their relationship with different types of social situations? How different emotions may get mixed and combined within successive combinatory levels? How prolonged sentiments and social moods may be established, detected, and changed upon entire communities? Some aspects of these questions will be discussed in the two Sections that follow.

4. Reference frames for social emotions

4.1. Representations of emotions

The discussion on a new frame of reference for these 'small' but frequent social emotions in daily life is one of the aims of our approach. To begin with a general scheme, the well-known Cartesian emotion diagram (POSNER et al., 2005; Russell, 1980; Russell and Barrett, 1999), the "circumplex model" counts with *valence* and *arousal* as the two fundamental dimensions of emotional space. See Fig. 2. This model has the advantage of placing several emotions in very appropriate places relative to each other. Actually, numerous graphical mappings of emotions have been derived from that model, often introducing a third dimension which usually is either *approach/avoidance* or *time*. The model may also provide a visual understanding of emotional trajectories when the valence and activation coordinates of subjects are changed according to the evolution of mental states. Further, the subject's persistence in some emotional state would be tantamount to a permanent displacement of the origin of coordinates so that more – or less – activation would be needed comparatively, or that a bigger – or smaller – increment would be needed regarding valence. The permanence in time of some emotional states is often considered as an instance of 'sentiment', at least in the way AI sentiment analysis is currently practiced, as will be discussed later.

Another relevant model to consider is due to Robert Plutchik (Plutchik, 1980), known as the "emotion wheel", with eight primary emotions grouped on positive versus negative influences and capable of combining to form emotional dyads and triads. This model has been followed by many professionals and counselors on personality disorders and self-improvement. But many other lists and classifications have also been developed, and the number of emotions listed has been steadily increasing as well as their possible combinations (Parrott, 2001). For instance, six axes of emotions each one with another six gradual ranges give a total of 36 emotions (Kort et al., 2001), and the Book of Human Emotions contains a total of 154 emotions and sentiments identifiable in different cultures (Watt Smith, 2021). It is interesting that Ekman's big six emotions were later complemented with another 16 by his research collaborators (Ekman, 1999), most of them social and not necessarily expressed in facial muscles.

Actually, one of the main problems of emotion theorists, at least for the research linked to commercial platforms, is not the compilation of possible emotion lists, but trying to establish a solid emotional underpinning for the most frequent social situations of daily life, either in front of a screen or in face-to-face relationships or in group contexts. Brute force approaches based on big data are useful for direct marketing purposes but not enough for developing a coherent perspective on emotions. As a revealing instance, laughter and crying, so basic emotional states in human close-knit groups and respectively fundamental for supporting the creation of social bonds and for mitigating their destruction (Marijuan and Navarro, 2010; Navarro et al., 2014, 2016), are still missing in almost all compilations of emotions. This absence tells us that relational aspects of importance might not be well solved yet in conventional approaches to emotion. The enigmatic role of laughter along all the stages of human life seems to consist in an indirect but highly efficient tool for bond-making via synaptic reinforcement (Navarro et al., 2016). Intriguingly, laughter appears as a "proto-phenomenon" of our ontogenetic sociality.

According to the previous sections, what natural intelligence would suggest us is the need of looking closely to the human life cycle (or better, the life course), searching for the systemic equivalents of cellular "sigma factors" and "checkpoints", it is to say, the fixed and flexible

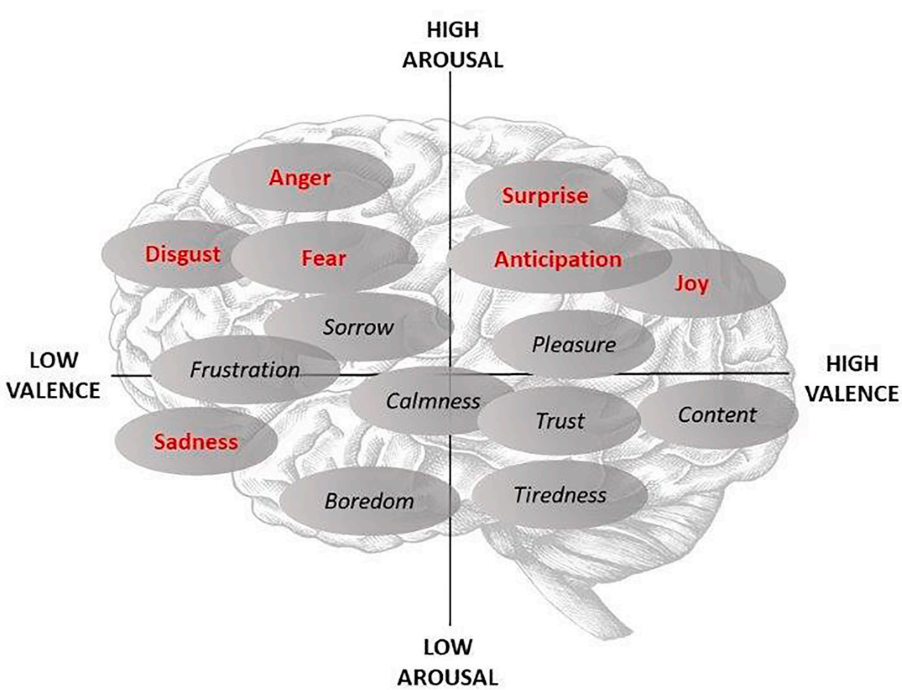

Fig. 2. The circumplex model: This Cartesian representation (arousal vs. valence) includes the eight basic/primary emotions mentioned in the main text (Section 5), in red, and another eight secondary emotions (in italics).

detection-patterns and the superstructural patterns already mentioned that are able to agitate our processing resources and reorient our cognitive mechanisms by means of emotional tools along the life course. An additional idea that the sociotype may convey is that our fundamental adaptation is to a very rich and structured social environment. So, we must try to synthesize what could be the main structural and superstructural patterns triggering our panoply of emotions in the adjustment to a very complex social environment—with remarkable differences along the life course, and with respect to gender, social location, cultures, etc.

4.2. Sociality different scenarios

Trying to separate different emotional domains pertaining to the main relational scenarios of our species, as sketched in the sociotype, we might distinguish the following modalities: (i) survival and self-maintenance; (ii) sex and family nucleus; (iii) friends, colleagues, and general acquaintances; and (iv) ultrasociality & collective identities.

Concerning the survival and self-maintenance drivers, (i) they would be served by the basic emotions, whatever number we may consider. (For instance, Ekman's six basic ones: sadness, happiness, fear, anger, surprise, and disgust.) And concerning sex, marriage, and family life, (ii) they would also be covered by these basic emotions but also by some new specific emotions linked to strong bonding processes, 'almost' exclusively human (for instance: love, affection, lust, play, laughter, curiosity). These six plus six could be considered as our *'primary'* emotions for strong bonding.

Then, we could put together another sector of the sociotype, 'friends', which is somehow intermediate between the nuclear relations and the work colleagues and general acquaintances, considering all of them together under the label of 'interindividual' (iii). In this interindividual domain we have the instinct to make social bonds of weaker nature, more numerous and malleable, implying frequent inclusions/exclusions. In this domain of weak bonds, we must maintain our reputation and personal image, we must cooperate to achieve our best interests with occasional conflicts with other individuals, and we instinctively abide by stringent relational rules, even at very early ages (Tomasello, 2019). This is the genuine territory where Trivers' moralistic emotions are deployed as spontaneous behavioral strategies in the reciprocity game (Pinker, 2009; Trivers, 1985). Another six *'secondary'* emotions would appear, say, in parallel to those primary emotions just mentioned for strong bonding. We would find for weak bonds: resentment, liking, gratitude, sympathy, guilt, and shame. In this point we should remind the important difference in social science between strong bonds and weak bonds, and the centrality of the latter in the establishment of commercial and economic activities (Granovetter, 1973). Weak bonds and their associated emotions would be the main support of civility. So, we can establish an interesting correspondence of bonds with emotions: primary and secondary emotions would respectively be in charge of creating and maintaining strong bonds and weak bonds.

In this interindividual scenario, we propose that in an analogy with the tridimensional approaches of the circumplex model, there would appear tree highly frequent distinctional or 'patterning axes' for the triggering of the emotions, precisely where most group conflicts arise. They would consist of: trust (cooperation) vs. mistrust (conflict); superiority (arrogance) vs. inferiority (humiliation); and inclusion (acceptance) vs. exclusion (rejection), which is highly significant in many cases where being marginalized in a group is tantamount to have really bleak a future. And an additional condition to consider relates to cognitive distance, familiar (close) vs. unfamiliar (distant), which is highly relevant concerning our 'automatic' minimization processing and the subsequent emotional response to the colligated patterns present in these previous axes (Collins, 1991; Collins and Marijuán, 1997). Thereafter, the respective coordinates of the different patterns in this multidimensional space, appropriately transformed by the 'familiar' vs. 'unfamiliar' condition, would call into action different emotions, either of strong bonding – primary – or the others of interindividual nature – secondary – more related to our enlarged prosocial inclination.

And there is also the very important ultrasociality phenomenon (iv). Historically, we have embarked in collective identities of highly variable nature and size. New determinants such as commonality (unity) vs. individualism (discord), freedom (tolerance) vs. oppression (intolerance), and equality (fairness, justice) vs. inequality (unfairness, injustice) represent further important dimensions or patterning axes to allocate our emotional responses regarding collective identities. It somehow echoes the classical political slogan of "*Liberté*, *Égalité*, *Fraternité*." But evolutionarily, the timing of history for this ultrasocial phenomenon has been too short. Quite probably, utrasociality has been using the emotional resources already present in Homo sapiens. So, it would have co-opted a mixing of basic and secondary emotions in the emergence and maintenance of the new social structures—particularly conveyed by means of political, religious, and cultural developments. However, like in the case of reading (Dehaene, 2009) it might well be that the socialization process of individuals, their 'education', provokes the emergence of genuine new emotions derived from the combinatorics among previous emotional reactions, with reactions such as: elation, admiration, awe, adhesion, synchronization, togetherness; as well as rejection, hostility, xenophobia …

In any event, our emotional minds are not organized in watertight compartments. The presence of the previous interindividual axes continues to be inevitable in this new ultrasocial domain too, as is the presence of the basic patterns and emotional reactions related to the close friends and family circle. So, it would be very frequent the overlapping and conflict between opposing emotional occurrences. Displaying these conflicts and showing how they can – or cannot – be solved is the bread and butter of many an artistic discipline (Booker, 2004), and the real substrate of personal wisdom in our social lives. It is in the ultrasocial domain where these conflicts appear more recurrently, given the easy recourse to primary emotions—the appeal to brotherhood or to hatred, or the use of fear as a means of mass control. The emotional mismatch between relational domains is currently amplified in social networks, where the inimical, offensive forms of primary emotions often substitute for the civilized relationships of weak bonds.

Summing up, this approach to social emotions based on natural intelligence and our social nature proposes a different kind of exploration. We do not think, at the time being, that a detailed listing of prosocial emotions taken in isolation would be feasible, or even interesting. Our alternative – complementary– approach based in the detection of patterning axes associated to the different relational domains, rather than being in opposition to Artificial Intelligence fields, could be developed in a fruitful cooperation with them. The basic ideas on Sentiment Analysis in next Section could provide some inkling on how to contribute meaningfully, in an empirical way, to refine and develop the present suggestions.

5. Emotions in AI: sentiment analysis

Emotions have never been alien in AI. Despite the first wave of pragmatic and rationalistic approaches, emotions and emotion-metrics were incorporated to AI relatively soon. From the rudimentary "sentometers" of Manfred Clynes in the 1980s (Clynes, 1988), representatives of the new wave to come, to the "affective computing" research headed by Rosalind Picard in the 1990s (Picard, 1997). Later on, with the growing business interest about the "attention economy", different emotion-detection systems and patents were issued by Facebook, Affectiva, Emoshape, and others, including the design of specific microchips (Zuboff, 2019). An important referent was Eksman's "big six" basic emotions, which became 16 in Facebook research. One of the main orientations was towards automatic visual detection of emotions in images as well as the identification of multiple emotional combinatorics. In another direction, Alex Pentland's "sociometers" were addressed to cover important metrics on social relationships: interactions, bonding processes, real hierarchies, etc.—a sociotype of sorts easily measurable via data from ad hoc wearables and mobiles (Pentland, 2014). Further, the boom of social networks, with their enormous trail of images, videos, and texts, has promoted new types of AI approaches. One of them, sentiment analysis, is the research link we take as a potential way to connect natural intelligence with artificial intelligence regarding social emotions (purposively, we leave aside the 'black box' approaches of machine learning, so fashionables in the new open AI language 'chat' systems).

Sentiment analysis is based on natural language processing. For each text of natural language (after its 'cleaning' and normalization) a global sentiment vector is processed, composed of multiple paragraphs. In each paragraph there is a count of the number of words associated with each basic emotion, taken from an ad hoc lexicon, obtaining the percentage of words associated with each emotion. The lexicon is a list of English words (or of other languages) with their associations with basic emotions (often eight basic emotions are considered: anger, fear, anticipation, trust, surprise, sadness, joy, and disgust). With each word there are two associated 'sentiments' (either negative or positive *valence*), or more values in a larger gradation. Next, from the values of the sentiment vector – the valence of the paragraphs – and the way emotions are distributed in the text, an overall statistical assessment may be obtained. Besides this emotional counting, there appears a valence or sentiment trajectory along the text, which is a plot of the variation of the emotional valence with respect to the narrative time. It is a trajectory graph quite revealing about the overall intentionality and mood of the text.

5.1. Relationship with natural intelligence views

Thereafter, looking for a fruitful interconnection of sentiment analysis methodologies with natural intelligence, what should matter is the possibility of detecting not just the emotions directly in collections of texts, the overall valence, or the trajectory plot, but to distil contextual patterns that may trigger the different emotions. It forms part of the big problem of 'context', plagued with semantic traps for whatever automated analysis. The six patterning axes we have previously identified, three for interindividual cases and another three for collective identities, could be tentatively approached via some of the new bootstrapping and subject classifier methods coupled with deep learning.

In order to facilitate this kind of AI exploration a parsimonious stance is needed. Perhaps, in the same way that "groups shape emotions" and "emotions shape groups" (van Kleef and Fischer, 2016), we may state that "contexts shape emotions" and "emotions shape contexts". This dialectics may be realized via sentiment analysis models endowed with adequate lexicons where emotion-laden words create contexts (which become identified as patterns), and the presence of the contextual patterns helps to recognize further emotional words. Lexicons should be built with more inner complexity, each word projected initially to all the (three or six) axes present, with a more complex valence gradation, and somehow making a reinforcing link with the other emotional words present in the sentence or paragraph. It is a bootstrapping methodology to be designed and implemented via NLP and sentiment analysis. To start with, simpler models related to each patterning-axis polarity could be developed.

In sum, the fast advancement of these new fields of sentiment analysis, opinion mining, affective computing, and emotional AI, as well as their multiple applications in social media, marketing, health care, social surveys, political forecasting, etc., although rather distant from the views herein advocated, could also represent a degree of opportunity for the advancement of new kinds of research on social emotions … Whether the future applications of AI may transcend the present "age of surveillance capitalism" (Zuboff, 2019) or not, may also depend on the relevance of the counterpoised research.

6. Concluding comments

Whatever the domain of life, it is the advancement of the life cycle what guides the informational interaction with the environment. We may consider in a general way that emotions are the inner forces that manipulate the ongoing behavioral (or genetico-molecular) trajectory of the biological system so that the new fitness opportunities in the environment may be properly realized along the life course.

Natural intelligence should study the organization of the "information flow" subtended with the environment, as well as the inner processing resources involved. The molecular tricks inherent in sigma factors, in eukaryotic checkpoints, or in the neuronal circuits of fixed, flexible, or superstructural perception-action patterns, become all of them ad hoc instances of (biological) natural intelligence.

In the approach to social emotions, we want to remark the research interest of the sociotype. Indeed, the genotype and phenotype constructs for the human species, notwithstanding their respective degrees of variability, could well be accompanied by a sociotype metrics, representing the relative constancy of the basic social environment to which the individuals of our species would be evolutionarily adapted.

The differentiated relational domains within the sociotype are important for a more nuanced approach to social emotions. We should emphasize our distinction between primary emotions for the closest relational circle ('strong bonds'), secondary emotions for the interindividual relationships ('weak bonds'), and finally some hypothetical tertiary emotions for the ultrasocial domain. Overall, there is design elegance, a manifest *economy*, in the evolutionary correspondence between groups of emotions and social bonding classes.

Then, as a research strategy, rather than looking for more sophisticate emotion classifications, or for more and more enlarged lists, we have pointed at the contextual patterns guiding our social-emotional adjustment. The interindividual and ultrasocial patterning axes herein proposed are just educated guesses, but via sentiment analysis bootstrap we could start to materialize the dialectical interrelationship contexts/emotions. Thereafter, the refining of classifications and the enlarging of emotion lists could be contemplated and developed in a new light.

Our final approach to sentiment analysis within current artificial intelligence has shown that there might be an effective and efficient connection between research programs of the two branches of intelligence—there is enough research potential to apply the muscle of AI to social emotions and their triggering circumstances. Natural intelligence versus artificial intelligence: what was separated should be reunited.

Data availability statements

The authors report no associated data.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research is partially funded (JNL) by 'Grupo Decisión Multicriterio Zaragoza' research group (S35-20R) and the project 'Participación Ciudadana Cognitiva y Decisiones Públicas. Aplicaciones Sociosanitarias' (Ref. LMP35_21), both supported by grants from the Regional Government of Aragon and FEDER funds; and also by the Project "Codecisión Cognitiva y Colaborativa. Aplicaciones en Salud" (Ref.: PID 2022-139863OB-I00), supported by the Ministerio de Ciencia e Innovación (Spain).

References

Allman, J., 1999. Evolving Brains. Scientific American Library, New York, NY.

Arbib, M.A., Fellous, J.-M., 2004. Emotions: from brain to robot. Trends Cognit. Sci. 8, 554–561. https://doi.org/10.1016/j.tics.2004.10.004.

Armitage, J.P., Holland, I.B., Jenal, U., Kenny, B., 2005. "Neural networks" in bacteria: making connections. J. Bacteriol. 187, 26–36. https://doi.org/10.1128/JB.187.1.26-36.2005.

Barrett, L.F., 2006. Solving the emotion paradox: categorization and the experience of emotion. Pers. Soc. Psychol. Rev. 10, 20–46. https://doi.org/10.1207/s15327957pspr1001_2.

Berry, E., De Geest, S., 2012. Tell me what you eat and I will tell you your sociotype: coping with diabesity. Rambam Maimonides Med J 3, e0010. https://doi.org/10.5041/RMMJ.10077.

Berry, E.M., 2011. The role of the sociotype in managing chronic disease: integrating bio-psycho-sociology with systems biology. Med. Hypotheses 77, 610–613. https://doi.org/10.1016/j.mehy.2011.06.046.

Booker, C., 2004. The Seven Basic Plots: Why We Tell Stories, Reprinted 2017. Bloomsbury Publishing PLC, London.

Borges, J.L., 1941. The Garden of Forking Paths, English 2018. Penguin Modern, New York.

Bryson, J.J., 2015. Artificial intelligence and pro-social behaviour. In: Collective Agency and Cooperation in Natural and Artificial Systems: Explanation, Implementation and Simulation, pp. 281–306. https://doi.org/10.1007/978-3-319-15515-9_15.

Calvo, P., 2016. The philosophy of plant neurobiology: a manifesto. Synthese 193, 1323–1343. https://doi.org/10.1007/s11229-016-1040-1.

Clynes, M., 1988. Generalised emotion how it may be produced, and sentic cycle therapy. In: Emotions and Psychopathology. Springer US, Boston, MA, pp. 107–170. https://doi.org/10.1007/978-1-4757-1987-1_6.

Collins, K.P., 1991. On the automation of knowledge within central nervous systems. In: Manuscript Presented at the AAAS Annual Meeting. Washington DC.

Collins, K.P., Marijuán, P.C., 1997. El Cerebro Dual: Un Acercamiento Interdisciplinar a la Naturaleza del Conocimiento Humano y Biológico. Editorial Hacer, Barcelona.

Corning, P.A., 2020. Beyond the modern synthesis: a framework for a more inclusive biological synthesis. Prog. Biophys. Mol. Biol. 153, 5–12. https://doi.org/10.1016/j.pbiomolbio.2020.02.002.

Csany, V., 1988. Contribution of the genetical and neural memory to animal intelligence. In: Jerison, H., Jerison, I. (Eds.), Intelligence and Evolutionary Biology. Springer-Verlag, Berlin, pp. 299–318.

Damasio, A., 1994. Descartes' Error: Emotion, Reason, and the Human Brain. Putnam, New York, NY.

Dehaene, S., 2009. Reading in the Brain. Penguin, New York, NY.

Dunbar, R.I.M., 2007. Groups, gossip, and the evolution of language. In: New Aspects of Human Ethology, pp. 77–89. https://doi.org/10.1007/978-0-585-34289-4_5.

Dunbar, R.I.M., 2004. The Human Story. Faber and Faber, London, UK.

Ekman, P., 1999. In: Dalgleish, T., Power, M. (Eds.), Handbook of Cognition and Emotion. John Wiley & Sons, Sussex.

Ekman, P., 1992. An argument for basic emotions. Cognit. Emot. 6, 169–200. https://doi.org/10.1080/02699939208411068.

Friston, K., 2012. A free energy principle for biological systems. Entropy 14, 2100–2121. https://doi.org/10.3390/e14112100.

Galperin, M.Y., 2005. A census of membrane-bound and intracellular signal transduction proteins in bacteria: bacterial IQ, extroverts and introverts. BMC Microbiol. 5, 35. https://doi.org/10.1186/1471-2180-5-35.

Gama-Castro, S., Salgado, H., Santos-Zavaleta, A., Ledezma-Tejeida, D., Muñiz-Rascado, L., García-Sotelo, J.S., Alquicira-Hernández, K., Martínez-Flores, I., Pannier, L., Castro-Mondragón, J.A., Medina-Rivera, A., Solano-Lira, H., Bonavides-Martínez, C., Pérez-Rueda, E., Alquicira-Hernández, S., Porrón-Sotelo, L., López-Fuentes, A., Hernández-Koutoucheva, A., Del Moral-Chavez, V., Rinaldi, F., Collado-Vides, J., 2016. RegulonDB version 9.0: high-level integration of gene regulation, coexpression, motif clustering and beyond. Nucleic Acids Res. 44, D133–D143. https://doi.org/10.1093/nar/gkv1156.

Gershenson, C., 2021. Intelligence as information processing: brains, swarms, and computers. Front Ecol Evol 9, 1–7. https://doi.org/10.3389/fevo.2021.755981.

Granovetter, M.S., 1973. The strength of weak ties. Am. J. Sociol. 78, 1360–1380. https://doi.org/10.1086/225469.

Grigoroudis, A.I., Panagiotidis, C.A., Lioliou, E.E., Vlassi, M., Kyriakidis, D.A., 2007. Molecular modeling and functional analysis of the AtoS–AtoC two-component signal transduction system of Escherichia coli. Biochim. Biophys. Acta Gen. Subj. 1770, 1248–1258. https://doi.org/10.1016/j.bbagen.2007.04.004.

Henrich, J., 2016. The Secret of Our Success. Princeton University Press, Princeton, NJ.

Igamberdiev, A.U., 2023. Reflexive structure of the conscious subject and the origin of language codes. Biosystems 231, 104983. https://doi.org/10.1016/j.biosystems.2023.104983.

Ji, S., 2017. Waves as the symmetry principle underlying cosmic, cell, and human languages. Information 8, 1–25. https://doi.org/10.3390/info8010024.

Kahneman, D., 2011. Thinking, Fast and Slow. Farrar, Straus and Giroux, New York.

Karp, P.D., Keseler, I.M., Shearer, A., Latendresse, M., Krummenacker, M., Paley, S.M., Paulsen, I., Collado-Vides, J., Gama-Castro, S., Peralta-Gil, M., Santos-Zavaleta, A., Peñaloza-Spínola, M.I., Bonavides-martinez, C., Ingraham, J., 2007. Multidimensional annotation of the Escherichia coli K-12 genome. Nucleic Acids Res. 35, 7577–7590. https://doi.org/10.1093/nar/gkm740.

Koole, S.L., 2009. The psychology of emotion regulation: an integrative review. Cognit. Emot. 23, 4–41. https://doi.org/10.1080/02699930802619031.

Kort, B., Reilly, R., Picard, R.W., 2001. An affective model of interplay between emotions and learning: reengineering educational pedagogy-building a learning companion. In: Proceedings - IEEE International Conference on Advanced Learning Technologies. ICALT, pp. 43–46. https://doi.org/10.1109/ICALT.2001.943850, 2001.

Lanham, R.A., 2006. The Economics of Attention: Style and Substance in the Age of Information. The University of Chicago Press, Chicago, IL.

Marijuán, P.C., 2002. Bioinformation: untangling the networks of life. Biosystems 64, 111–118. https://doi.org/10.1016/S0303-2647(01)00179-4.

Marijuán, P.C., 1989. Natural Intelligence: the Evolution of Biological Information Processing. University of Barcelona (PhD Thesis), Barcelona.

Marijuán, P.C., del Moral, R., Ji, S., Lacruz, M.G., Gómez-Quintero, J.D., Navarro, J., 2019. Fundamental, quantitative traits of the "sociotype.". Biosystems 180, 79–87. https://doi.org/10.1016/j.biosystems.2019.02.007.

Marijuán, P.C., Montero-Marín, J., Navarro, J., García-Campayo, J., del Moral, R., 2017. The "sociotype" construct: gauging the structure and dynamics of human sociality. PLoS One 12, e0189568. https://doi.org/10.1371/journal.pone.0189568.

Marijuán, P.C., Navarro, J., 2022. The biological information flow: from cell theory to a new evolutionary synthesis. Biosystems 213, 104631. https://doi.org/10.1016/j.biosystems.2022.104631.

Marijuan, P.C., Navarro, J., 2010. The Bonds of Laughter: A Multidisciplinary Inquiry into the Information Processes of Human Laughter. http://arxiv.org/abs/1010.5602.

Marijuán, P.C., Navarro, J., del Moral, R., 2018. How prokaryotes 'encode' their environment: systemic tools for organizing the information flow. Biosystems 164, 26–38. https://doi.org/10.1016/j.biosystems.2017.10.002.

Marijuán, P.C., Navarro, J., del Moral, R., 2010. On prokaryotic intelligence: strategies for sensing the environment. Biosystems 99, 94–103. https://doi.org/10.1016/j.biosystems.2009.09.004.

McCarthy, J., Misnky, M.L., Rochester, N., Shannon, C.E., 2006. A proposal for the dartmouth summer research project on artificial intelligence, 1955 AI Mag. 27 (no.4), 12–14.

Minelli, A., 2015. Grand challenges in evolutionary developmental biology. Front Ecol Evol 2 (JAN), 1–11. https://doi.org/10.3389/fevo.2014.00085.

Navarro, J., Cañete, M., Olivera, F.J., Gil-Lacruz, M., Gil-Lacruz, A., Marijuán, P.C., 2022. The cost of loneliness: assessing the social relationships of the elderly via an abbreviated sociotype questionnaire for inside and outside the clinic. Int. J. Environ. Res. Publ. Health 19, 1253. https://doi.org/10.3390/ijerph19031253.

Navarro, J., del Moral, R., Alonso, M.F., Loste, P., Garcia-Campayo, J., Lahoz-Beltra, R., Marijuán, P.C., 2014. Validation of laughter for diagnosis and evaluation of depression. J. Affect. Disord. 160, 43–49. https://doi.org/10.1016/j.jad.2014.02.035.

Navarro, J., del Moral, R., Marijuán, P.C., 2016. Laughing bonds. Kybernetes 45, 1292–1307. https://doi.org/10.1108/K-02-2016-0026.

Navarro, J., Marijuán, P.C., 2022. The natural, artificial, and social domains of intelligence: a triune approach. In: IS4SI 2021. MDPI. https://doi.org/10.3390/proceedings2022081002.

Panksepp, J., 2015. Toward the constitution of emotional feelings: synergistic lessons from izard's differential emotions theory and affective neuroscience. Emotion Review 7, 110–115. https://doi.org/10.1177/1754073914554788.

Parrott, W.G., 2001. Emotions in Social Psychology: Essential Readings. Edwards, Ann Arbor.

Pentland, A., 2014. Social Physics: How Good Ideas Spread-The Lessons from a New Science. The Penguin Press, New York, NY.

Perez Velazquez, J.L., 2009. Finding simplicity in complexity: general principles of biological and nonbiological organization. J. Biol. Phys. 35, 209–221. https://doi.org/10.1007/s10867-009-9146-z.

Picard, R., 1997. Affective Computing. MIT Press, Cambridge, MA.

Pinker, S., 2009. How the Mind Works. W. W. Norton, New York, NY.

Plutchik, R., 1980. A general psychoevolutionary theory of emotion. In: Theories of Emotion. Elsevier, pp. 3–33. https://doi.org/10.1016/B978-0-12-558701-3.50007-7.

Posner, J., Russell, J.A., Peterson, B.S., 2005. The circumplex model of affect: an integrative approach to affective neuroscience, cognitive development, and psychopathology. Dev. Psychopathol. 17 (3), 715–734. https://doi.org/10.1017/S0954579405050340.

Ramón y Cajal, S., 1899. Textura del Sistema Nervioso del Hombre y de los Vertebrados. Nicolás Moya, Madrid.

Russell, J.A., 1980. A circumplex model of affect. J. Pers. Soc. Psychol. 39 (6), 1161–1178. https://doi.org/10.1037/h0077714.

Russell, J.A., Barrett, L.F., 1999. Core affect, prototypical emotional episodes, and other things called emotion: dissecting the elephant. J. Pers. Soc. Psychol. 76, 805–819. https://doi.org/10.1037/0022-3514.76.5.805.

Salgado, H., Peralta-Gil, M., Gama-Castro, S., Santos-Zavaleta, A., Muñiz-Rascado, L., García-Sotelo, J.S., Weiss, V., Solano-Lira, H., Martínez-Flores, I., Medina-Rivera, A., Salgado-Osorio, G., Alquicira-Hernández, S., Alquicira-Hernández, K., López-Fuentes, A., Porrón-Sotelo, L., Huerta, A.M., Bonavides-Martínez, C., Balderas-Martínez, Y.I., Pannier, L., Olvera, M., Labastida, A., Jiménez-Jacinto, V., Vega-Alvarado, L., del Moral-Chávez, V., Hernández-Alvarez, A., Morett, E., Collado-Vides, J., 2013. RegulonDB v8.0: omics data sets, evolutionary conservation, regulatory phrases, cross-validated gold standards and more. Nucleic Acids Res. 41, D203–D213. https://doi.org/10.1093/nar/gks1201.

Shklovskiy-Kordi, N.E., Igamberdiev, A.U., 2022. Natural computation and its limits: efim Liberman at the dawn of a new science. Biosystems 215–216, 104653. https://doi.org/10.1016/j.biosystems.2022.104653.

Silk, J.B., 2007. Social components of fitness in primate groups. Science 317 (5843), 1347–1351. https://doi.org/10.1126/science.1140734.

Slijepcevic, P., 2018. Evolutionary epistemology: reviewing and reviving with new data the research programme for distributed biological intelligence. Biosystems 163, 23–35. https://doi.org/10.1016/j.biosystems.2017.11.008.

Solé, R., Moses, M., Forrest, S., 2019. Liquid brains, solid brains. Phil. Trans. Biol. Sci. 374, 20190040 https://doi.org/10.1098/rstb.2019.0040.
Timsit, Y., Grégoire, S.-P., 2021. Towards the idea of molecular brains. Int. J. Mol. Sci. 22, 11868 https://doi.org/10.3390/ijms222111868.
Tomasello, M., 2019. Becoming Human. Harvard University Press, Cambridge.
Tozzi, A., Peters, J.F., Fingelkurts, Andrew A., Fingelkurts, Alexander A., Marijuán, P.C., 2017. Topodynamics of metastable brains. Phys. Life Rev. 21, 1–20. https://doi.org/10.1016/j.plrev.2017.03.001.
Trewavas, A., 2017. The foundations of plant intelligence. Interface Focus 7, 20160098. https://doi.org/10.1098/rsfs.2016.0098.
Trivers, R., 1985. Social Evolution. Benjamin-Cummings Publishing Co, San Francisco, CA.
Turchin, P., 2016. Ages of Discord. Beresta Books LLC, Chaplin, CT.
Turón, A., Altuzarra, A., Moreno-Jiménez, J.M., Navarro, J., 2023. Evolution of social mood in Spain throughout the COVID-19 vaccination process: a machine learning approach to tweets analysis. Publ. Health 215, 83–90. https://doi.org/10.1016/j.puhe.2022.12.003.
Ulrich, L.E., Koonin, E.V., Zhulin, I.B., 2005. One-component systems dominate signal transduction in prokaryotes. Trends Microbiol. 13, 52–56. https://doi.org/10.1016/j.tim.2004.12.006.
van Duijn, M., 2017. Phylogenetic origins of biological cognition: convergent patterns in the early evolution of learning. Interface Focus 7, 20160158. https://doi.org/10.1098/rsfs.2016.0158.
van Kleef, G.A., Fischer, A.H., 2016. Emotional collectives: how groups shape emotions and emotions shape groups. Cognit. Emot. 30, 3–19. https://doi.org/10.1080/02699931.2015.1081349.
Watt Smith, T., 2021. The Book of Human Emotions: an Encyclopedia of Feeling from Anger to Wanderlust [WWW Document]. https://web.archive.org/web/20210418032350/. http://anarchiveforemotions.com/files/DisOrder_uploads/images/TheBookOfHumanEmotions.pdf.
Whiten, A., Byrne, R.W., 1989. Machiavellian Intelligence: Social Expertise and the Evolution of Intellect in Monkeys, Apes, and Humans. Oxford Science Publications, Oxford.
Whiten, A., Byrne, R.W., 1988. Tactical deception in primates. Behav. Brain Sci. 11, 233–244. https://doi.org/10.1017/S0140525X00049682.
Wurtz, T., 2021. Nested information processing in the living world. Ann. N. Y. Acad. Sci. 1500, 5–16. https://doi.org/10.1111/nyas.14612.
Zuboff, S., 2019. Surveillance capitalism and the challenge of collective action. New Labor Forum 28, 10–29. https://doi.org/10.1177/1095796018819461.

Public Health 215 (2023) 83–90

Contents lists available at ScienceDirect

Public Health

journal homepage: www.elsevier.com/locate/puhe

Original Research

Evolution of social mood in Spain throughout the COVID-19 vaccination process: a machine learning approach to tweets analysis

A. Turón, A. Altuzarra, J.M. Moreno-Jiménez, J. Navarro*

Grupo Decisión Multicriterio Zaragoza (GDMZ), Dpt. Economía Aplicada, Facultad de Economía y Empresa, Universidad de Zaragoza, Gran Vía 2, 50003, Zaragoza, Spain

ARTICLE INFO

Article history:
Received 18 July 2022
Received in revised form
5 December 2022
Accepted 8 December 2022
Available online 14 December 2022

Keywords:
COVID-19 vaccination process
Sentiment analysis
Machine learning
Multivariate statistics
Tweets
Social mood

ABSTRACT

Objectives: This paper presents a new approach based on the combination of machine learning techniques, in particular, sentiment analysis using lexicons, and multivariate statistical methods to assess the evolution of social mood through the COVID-19 vaccination process in Spain.
Methods: Analysing 41,669 Spanish tweets posted between 27 February 2020 and 31 December 2021, different sentiments were assessed using a list of Spanish words and their associations with eight basic emotions (anger, fear, anticipation, trust, surprise, sadness, joy and disgust) and three valences (neutral, negative and positive). How the different subjective emotions were distributed across the tweets was determined using several descriptive statistics; a trajectory plot representing the emotional valence vs narrative time was also included.
Results: The results achieved are highly illustrative of the social mood of citizens, registering the different emerging opinion clusters, gauging public states of mind via the collective valence, and detecting the prevalence of different emotions in the successive phases of the vaccination process.
Conclusions: The present combination in formal models of objective and subjective information would therefore provide a more accurate vision of social reality, in this case regarding the COVID-19 vaccination process in Spain, which will enable a more effective resolution of problems.

Introduction

The COVID-19 outbreak has been declared a pandemic by the World Health Organization because of its high rate of spread, severity and its frequent outcomes of severe pneumonia, respiratory failure and death.[1] Vaccination has become the main available public resource against the pandemic. However, the prejudices or sentiments of the general public and political leaders, as reflected in social media, are having a significant impact on the progression towards achieving vaccination targets.[1,2]

Social media such as Twitter, Facebook, YouTube and LinkedIn, with billions of users worldwide,[3] represent the preferred sites for sharing, almost instantly and very easily, thoughts, feelings and opinions on all kinds of events.[4] Twitter[5] is one of the most active platforms with approximately 290.5 million monthly active users worldwide in 2020 and was projected to keep increasing up to over 340 million users by 2024.[6] Every second around 6000 tweets on average are tweeted, which corresponds to more than 350,000 tweets sent per minute, 500 million tweets per day and around 200 billion tweets per year.[7]

Tweets are real-time messages with a maximum length of 280 characters at a time. They can be analysed based on *hashtags*, which refer to the symbol (#) in Twitter (for instance: #COVID19), containing a combination of the word *hash* from ‘hash mark’ and the word *tag*, that marks something belonging to a specific category. Hashtags make it easy to quickly find messages about a topic of interest as well as to collect all the sentiments and opinions of people in one place or country.[8–11]

One of the most promising methods for content analysis in social media is sentiment analysis.[12,13] It can be understood as a set of approaches, techniques and tools that extracts people's opinions, feelings and thoughts from users' text data by means of natural language processing methods.[14] Sentiment analysis through social media is growing rapidly within the international scientific community as a useful tool to understand people's opinions and attitudes on any important situation or phenomenon that affects public opinion.[11,15] For instance, natural disasters,[11] the Syrian refugee crisis,[4] the UK-EU referendum,[16] the impact of Brexit,[17] presidential

* Corresponding author.
E-mail address: jnavarrol@unizar.es (J. Navarro).

https://doi.org/10.1016/j.puhe.2022.12.003

or general elections in the United States,[18,19] Indonesia[20] and India,[21] the world cup soccer tournament,[22] extremism in social media,[23] 2019 EVALI outbreak[24] and the COVID-19 outbreak.[25,26]

This article presents a new approach based on the combination of machine learning techniques, in particular, sentiment analysis using lexicons, and multivariate statistical methods to assess the evolution of social mood through the COVID-19 vaccination process in Spain via tweet messages. Sentiment analysis, or opinion mining, will allow us to carry out the quantitative scrutiny of those tweets by extracting subjective information from the detection of eight basic emotions (anger, fear, anticipation, trust, surprise, sadness, joy and disgust) and the assessment of polarity (valence), that is, the neutral, positive or negative connotation of the language used. Multivariate statistical methods, or data mining, will provide figures and graphics that can synthesise objective information and knowledge about the vaccination process; in particular, properties of social structures and the patterns of relationships among actors.

The proposed methodology has been applied to the analysis of 41,669 tweets from February 2020 to December 2021. It shows how the opinions expressed in social media can be analysed, so that the social mood of citizens can be detected, opinion groups and their leaders can be identified, and social support for government measures can be evaluated.[27–30] The present combination in formal models of objective and subjective information about the vaccination process provides a more accurate vision of reality, which will enable a more effective resolution of problems.

Vaccination process in Spain

The vaccination strategy in Spain was published on 2 December 2020, with 11 updates up to the end of the considered period for analysis.[31] Four phases were defined according to available doses (see Table 1). The population groups to be vaccinated were established in order of priority , following an assessment based on criteria that incorporated the risk of exposure and transmission, the existence of previous serious illness, and the socio-economic impact of the pandemic on each population group.[32]

Methods

The methodological approach was based on Social Web Mining complemented with natural language processing and social network analysis. Messages were collected from social networks, preprocessed, and then their features were extracted to perform an analysis of society's opinion and mood regarding that critical event, and the way people related to each other and exchanged information on that event on social networks. The chart in Fig. 1 shows the methodological procedure that consists of three steps and three stages for each step.

Step 1: Corpus Determination

Stage 1.1. Data collection

We used a data set of 300,286 tweets in Spanish, posted between 27 February 2020 and 31 December 2021, that is, from the beginning of the pandemic until the end of the main stage of the vaccination process in Spain. The tweets were extracted from Twitter using the *twitterR* package, written in R programming language, accessing Twitter API 2.0. and searching in the full historical Twitter database. The search key was built from the following hashtags: #covid; #covid19; #*Yomevacuno* (I'm getting vaccinated); #*Yonomevacuno* (I'm not getting vaccinated); #*Negacionista* (denialist). The key string used to query the database was (*covid* OR *covid19*) AND (*Yomevacuno* OR *Yonomevacuno* OR *negacionista*).

It was referring to COVID and vaccination and to the pro- and anti-vaccine positions. The search terms were written in Spanish, and the condition that the messages be written in Spanish was added.

The attributes extracted from each tweet and its author were stored in two separate tables in the database according to the scheme shown in Table 2.

Other R packages such as *httr*, *RCurl* or *jsonlite* were used to extract the information from the Twitter API, in addition to *RMySQL* to manage the data through a MySQL database.

Stage 1.2. Data preprocessing

The tweets were preprocessed to eliminate all elements of the data that are susceptible to inconsistency or ambiguity, or, for reasons of efficiency, unnecessary in the subsequent analysis (punctuation marks, symbols or numbers, and words that do not provide meaning). This means that from a total of 7,377,533 words, 5,813,263 were preserved after the depuration; in other words, 21.20% of the words were suppressed. The preprocessing was carried out using the *stringr* R package.

Table 1
Spanish vaccination phases according to available doses.

Phase/description	Duration	Population group
Phase 0/Development, authorisation and evaluation	From February 27 till 18 December 2020 (1st update)	
Phase 1/First available doses	From 19 December 2020 till 26 February 2021 (4th update)	• Residents and staff in nursing homes and care centres for the elderly and the highly dependent • Front-line health and social personnel • Other health and social care staff • Non-institutionalised major dependents
Phase 2/More available doses	From 27 February 2021 till 11 May 2021 (7th update)	• Over 80 years • People between 70 and 79 and people with very high-risk conditions • People between 60 and 65 • People between 66 and 69 • Other health and social care workers • Workers with an essential social function • People between 50 and 59
Phase 3/Widely available vaccine	From 12th May 2021 till 31st December 2021	• People between 40 and 49 • People between 30 and 39 • People between 20 and 29 • People between 12 and 19 • People between 5 and 11 • Booster doses

Table 2
Structure of the database.

Tweet		Author	
Tweet ID	Text	Author ID	Registration date
Author ID	Hashtags	Author name	Location
Creation date	Is retweeted	Username	Description

Table 3
Filters for corpus determination.

Filter	Number of tweets
Tweets collected	300,286
Tweets containing location	188,392
Authors geolocated in Spain	28,285
Authors geolocated in Spain with indication of region	24,394
Tweets posted by authors geolocated in Spain	41,669

Stage 1.3. Geolocation of the authors

To select the tweets written by Spanish authors, the geographical location of the authors was identified, when possible, from the information contained in the location field. This was done by calling the *Nominatim* geocoding service, an Open Data project/of *OpenStreetMap*.[33] A total of 188,392 tweets were posted by authors that contained information in this field, of which Nominatim obtained a location determined by its latitude, longitude and country. It was shown that 28,285 authors were from Spain and writing in Spanish, of which 24,394 had indication of the region.

The study considered the tweets sent by these 28,285 Spanish authors. In total, there were 41,669 tweets that constituted the corpus of the study, being some of them retweets of other authors (Table 3).

Step 2: Social mood evolution

Stage 2.1: Social network analysis

The most relevant network interaction was considered to be the retweet because the number of retweets was very abundant in the corpus and the action of sharing or retweeting a text implied personal interest from the person who retweeted. Given the list of 28,285 Spanish users, all their messages that were retweets were selected, and the authors of the original message were extracted

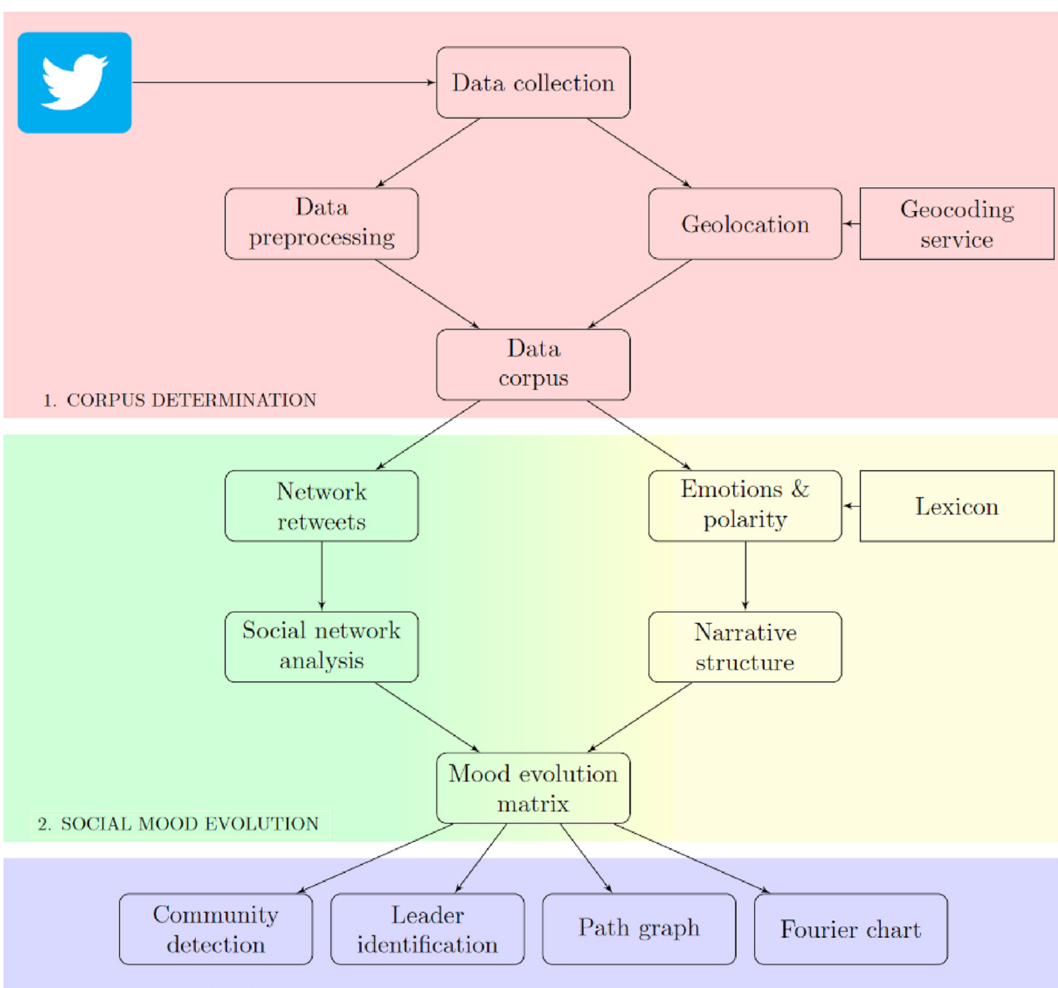

Fig. 1. Methodology flow diagram for the study of social mood evolution.

(although these may not be geolocated in Spain). A network was created based on the following methodological considerations:

- The network was a directed graph, the origin of each arc was the node corresponding to the author who retweeted a message and destination was the node that represented the author of the original tweet.
- The nodes were the users who had published tweets and retweets.
- The size of the nodes was proportional to the in-degree, representing the volume of retweets that has been made of their tweets.
- The colours of the nodes indicate communities. These communities have been calculated with the *Gephi software*,[34] which uses the algorithm described in.[35]
- The colour of the edges is the same as in the origin node, whereas their size is proportional to the number of messages from the destination node that the origin node has retweeted.
- The position of each node in the graph has been calculated using the *Force Atlas 2* algorithm,[36] an energy model for network spatialisation so that the more retweets a node has, the more focused it will be with respect to the nodes connected to it.

The resulting network contained 10,021 nodes and 17,340 edges, which represents a very low density, practically zero. Also, the average degree of the network is 1.73. This means that few retweets were made, and usually, the same authors were retweeted.

The analysis reveals the most influential users because of the size of their node (number of times a message of theirs has been retweeted) and their position within the cluster to which they belong (the more focused, the larger this size is). And the more compact a community is, the more relationships appear between its members. On the other hand, the different communities are closer to each other depending on how many nodes of each one are related to the other. The more relationships there are between two communities, the closer they would be.

Stage 2.2: Sentiment analysis

The 41,669 tweets were analysed, applying text mining by means of the *Syuzhet* 1.0.6 package[37] and *RStudio* 1.1.419, according to the general procedure already shown in Fig. 1.

As a first step, the sentiment was evaluated with *NRC Word-Emotion Association Lexicon Version 0.92*.[38–40] This lexicon is a list of English words and their associations with eight basic emotions (anger, fear, anticipation, trust, surprise, sadness, joy and disgust) and two sentiments (negative and positive). For each tweet, the valence was also obtained, that is, the difference between the number of positive and negative words, as well as the number of words associated with each of the above emotions and sentiments. We then examined how emotions were distributed throughout the text. To do this, several descriptive statistics were obtained (minimum, maximum, Q1, Q3, mean, and median) with which an overall assessment of each tweet could be achieved.

Stage 2.3: Mood evolution Matrix

After performing the social network and sentiment analysis (Stage 2.2 and Stage 2.3), the result is a matrix where the rows are the different tweets (41,669) and the columns (40) are grouped into the following information blocks:

- Tweet variables (8 columns): id, author_id, date, text, clean text, hashtag, retweeted (yes or no), retweeted_id.
- User variables (14 columns): name, username, created_at, location, description, type, lat, lng, country, city, region, postal code, cod_region, id_region.
- Emotions (8 columns): eight basic emotions (anger, fear, anticipation, trust, surprise, sadness, joy and disgust).
- Sentiments (4 columns): polarity (negative or positive), valence and number of sentiment words.
- Statistics (6 columns): six descriptive statistics (min, max, Q1, Q3, mean and median).

Results

This section presents the results corresponding to Step 3 of the methodology (Graphic Visualization). It includes illustrations of community detection, leader identification and path and Fourier graphs.

Community detection

Fig. 2 analyses the evolution of the retweet network during the phases of the process.

The most striking result is that two differentiated nuclei emerged, with very few interconnections between them, are distinguished in each phase: on the left, groups linked to the official sources of the Government and the health administrations of Spain, journalists and media (provaccine messages); on the right, accounts disseminating denialist and antivaccine messages. In Phase 0, there were 3818 users (2746 pro- and 1072 anti-vaccination); in Phase 1, 7758 users (5726 pro- and 2032 anti-vaccination); in Phase 2, 3510 users (2883 pro- and 627 anti-vaccination); and in Phase 3, 5637 users (2698 pro- and 2939 anti-vaccination). The composition and size of both pro- and anti-vaccine groups are clearly related to the variations produced in the social mood that will appear later in Fig. 3.

Leader identification

As can be seen in Table 4, there were several leaders involved in the different communities.

To better identify the leaders of the different communities, @sanidadgob corresponds to the official account of the Spanish Ministry of Health; @We_T_Resistance is an account positioned against the vaccination process; @salvadorilla (at the time Minister of Health of Spain); @rimbaudarth is an account positioned with the thesis of @We_T_Resistance; @publico_es is a media positioned in favour of the process; @Javier_CB is a very heterogeneous community with media presence but with very low activity on the network; and @daandina is a facultative working in public health. Clearly, the two most prominent leaders are the Government (1581 retweets) and the deniers (992 retweets).

Path and fourier graphs

The protocol described in Sections 3 and 4 (and Fig. 1) was applied to the 41,669 tweets. Fig. 3 shows the Fourier plot trajectory that represents emotional valence vs percentage of tweets (tweets date). From this analysis of tweets, we can see how the mental state or social mood of Spanish people has been changing through the different phases of the vaccination process (in different colours).

As shown in Fig. 3, the highest value of valence is found at Phase 1 (orange), between 4 and 6 January 2021, corresponding with the start of vaccination in Spain with Pfizer-BioNTech COVID-19 vaccine and the approval of Moderna COVID-19 (MD) vaccine by the European Medicines Agency. While the lowest value of valence is found at Phase 3 (green), between 4 and 6 August 2021, corresponding with the announcement of the need for booster doses and the debate on compulsory vaccination. On the other hand, we should note that the biggest fluctuations were produced in Phase 2

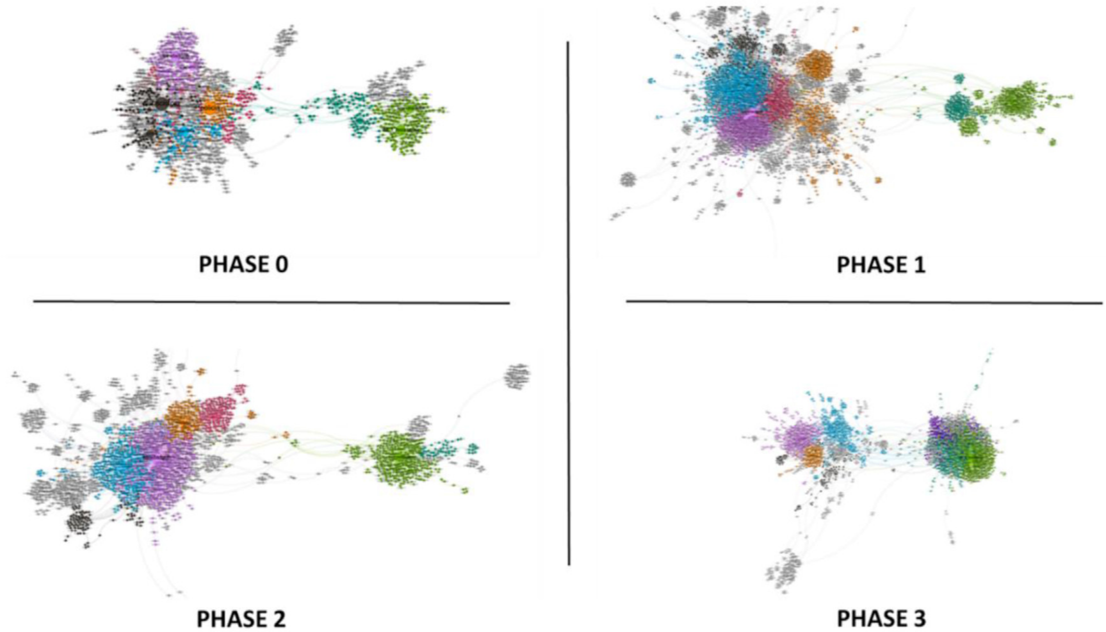

Fig. 2. Retweets network of the vaccination phases. The nodes are the users, and the arcs point goes from the retweeter to the author of the original tweet. The most retweeted authors are highlighted, and seven relatively clear clusters can be distinguished (each of them is formed by more than 2.5% of the total nodes and coloured in different colours). Within each cluster, those with highest number of retweets have been distinguished, appearing as the largest nodes in the graph.

(yellow) and Phase 3 (green) because of discordant health decisions on the Astra Zeneca vaccine.

Fig. 4 shows the percentage of words for each emotion according to each of the phases. It shows that the highest values for the main two emotions of the population during COVID-19 (fear and sadness)[41] were found at Phases 0 and 3. However, the highest value of joy and trust (more positive emotions) were shown in Phases 1 and 2, coinciding with the results obtained in Fig. 3 where the positive valences were in Phases 1 and 2.

The same pattern can be observed in Fig. 5 where we analyse the percentage of words for each phase according to each of the emotions. It is worth noting that the highest percentages of words

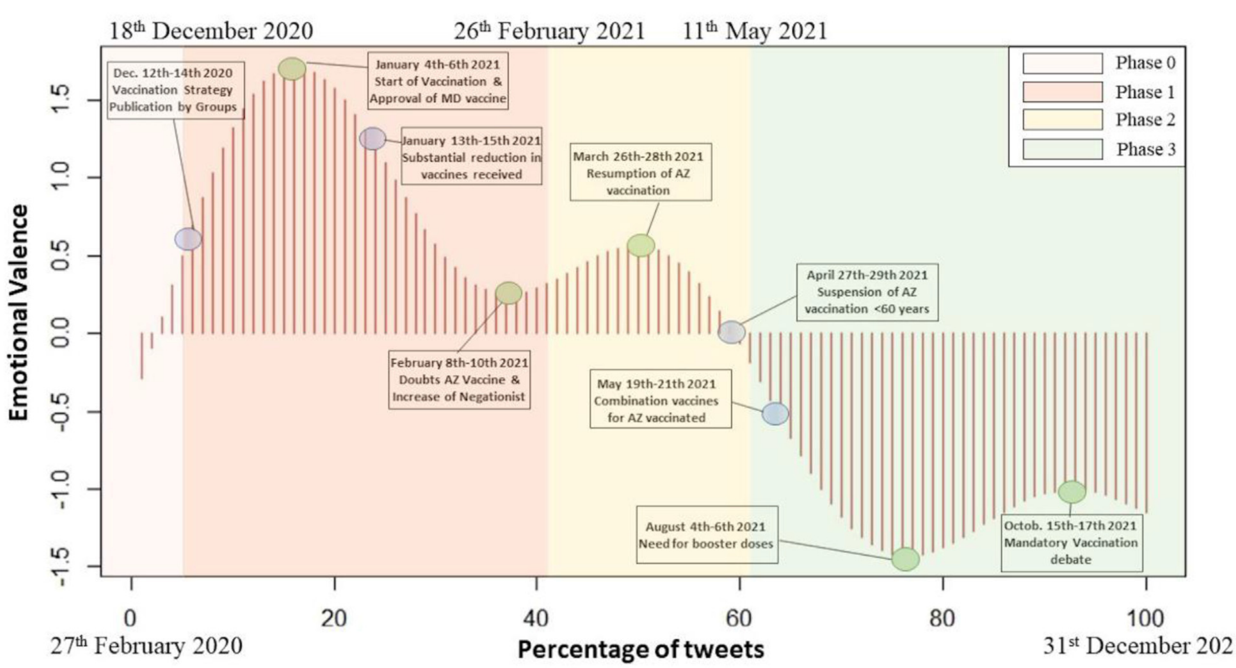

Fig. 3. Fourier plot trajectory of the tweets with the four phases (differently coloured). It represents emotional valence vs percentage of tweets (tweets date). In the upper side, the positive sentiments, and in the lower side, the negative ones. Local hotspots (green circles) and areas of trend change (purple circles) were marked by analysing the content of these tweets and relating them to relevant news and political decisions.

Table 4
Most retweeted authors in each community.

Community	Number of members (%)	Username	Number of retweets	Number of retweets (community)
Pink	1508 (15.05%)	@sanidadgob	1581	45.76%
Orange	446 (4.45%)	@daandina	42	6.87%
Black	774 (7.725)	@salvadorilla	347	17.08%
Fuchsia	426 (4.25%)	@Javier__CB	85	13.78%
Blue	1162 (11.60%)	@publico_es	158	7.52%
Green	1811 (18.07%)	@We_T_Resistance	922	23.28%
Emerald	292 (2.91%)	@rimbaudarth	171	34.76%

expressing the most negative emotions (anger, disgust, fear and sadness) are found in Phase 3, where the vaccines were widely available, but nevertheless, many doubts arose about the vaccination process with the news of the need for new doses or even compulsory vaccination. On the other hand, the most positive emotions (trust and joy) were in Phase 1, coinciding with the first available doses and the start of the vaccination process in Spain.

Discussion

This study has obtained a series of congruent results regarding the social networks involved, the evolution of social mood coupled with the dynamics of these networks, and the sentiment analysis represented in the plot trajectory. This overall congruence between the different kinds of obtained results may be interpreted as a very promising aspect of the approach.

Let us first point out that, regarding the evolution of social networks depicted in Fig. 2, the clustering dynamics during the four phases distinguished is surprisingly accurate, capturing the evolution of public opinion during the vaccination process. The analysis of the network of retweets not only shows the interconnections and clustering of the community of tweeters around interest groups but also shows how the structure of these groups varies throughout the process. It can be seen how public health decisions and other environmental circumstances that cause the changes in mood are translated not only into how tweeters are grouped but also who their referents are when it comes to sharing information. In addition, we can see in the network dynamics that clustering around two compact groups, of pro-vaccines and anti-vaccines, polarises the position of individuals in two communities with extremely few interconnections. These 'radical' divisions occur because of, and are exacerbated by, increasing conflict in communications about contentious topics such as lockdowns and compulsory vaccination.

Table 4 indicates the importance of public health communication from official sources (@sanidadgob and @salvadorilla) because their retweets from other users can reach far more people that are not following the official accounts. This means a cost-effective communication strategy for public health promotion.[42] In this regard, we may realise that most international political leaders are progressively turning to social networks to broadcast information about the pandemics, response plans, public health measures and connection with citizens.[43] This implies a series of strategic choices to use a more positive frame to influence opinion and action and to encourage compliance with public health norms and standards. The choice of positive frames may guide the national conversation away from seeking 'blame' for the pandemic towards a supportive mood necessary to implement the public health strategies required.[44] Finally, identifying and monitoring those social leaders whose opinions most closely reflect the needs or demands of society will contribute to make more realistic and effective public health decisions.

The prevalence of the different emotions during each of the phases shown in Figs. 4 and 5 would correlate well with the above. The high levels of anger, disgust, fear and sadness in phase 3 would document, as already said, the news about the new doses needed and the compulsory vaccination. The mental fatigue after the prolonged lockdowns and the stress for such long periods of uncertainty and pandemic fears are indeed reflected in the emotional arousal seen in these final phases.

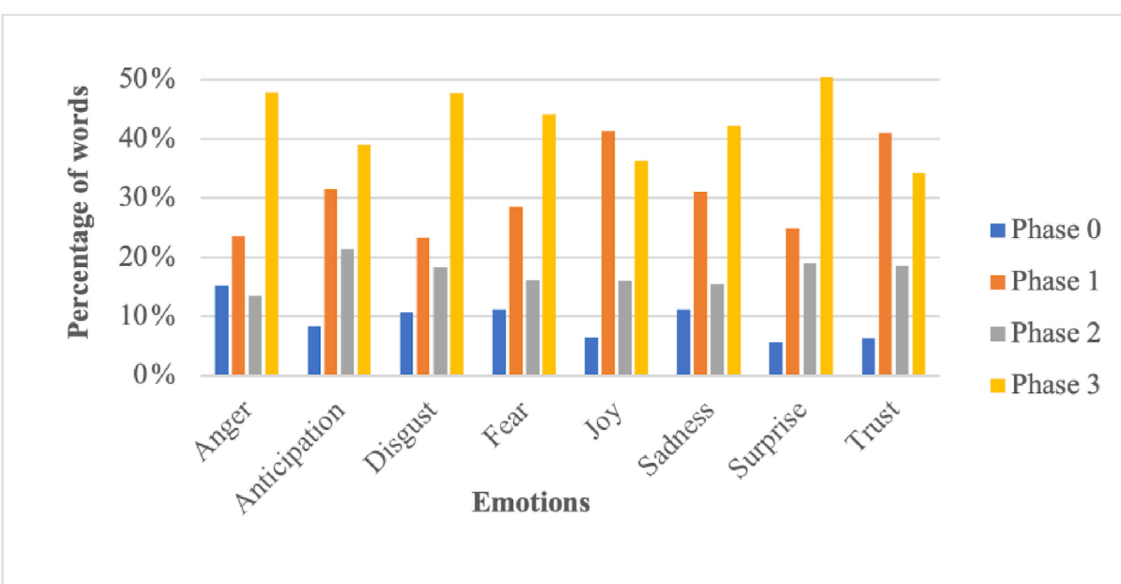

Fig. 4. Percentage of words per emotion according to each of the phases.

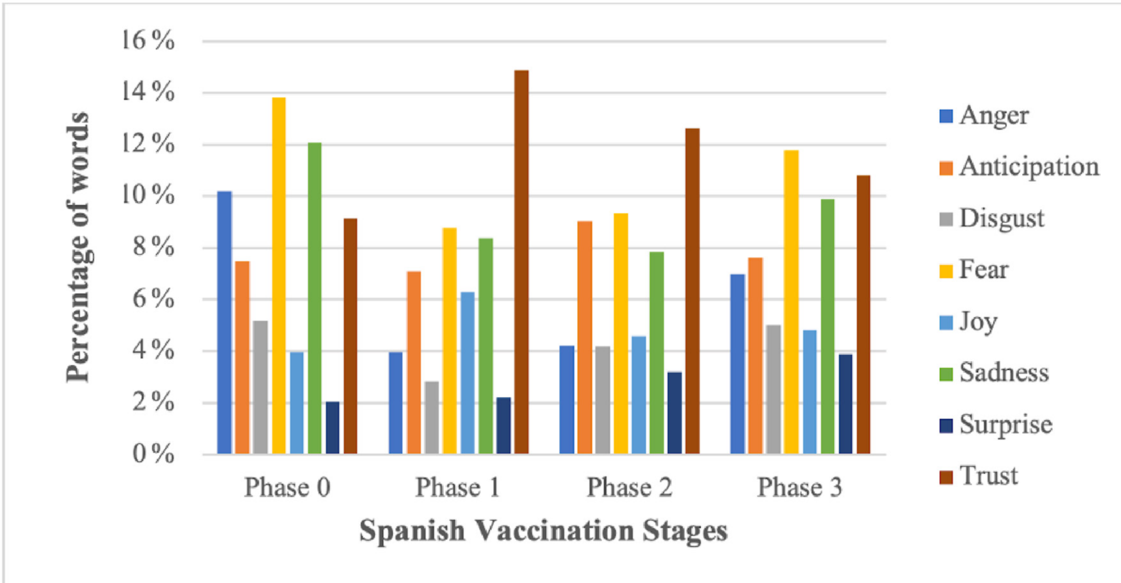

Fig. 5. Percentage of words per phase according to each of the emotions.

The specific results of sentiment analysis in the Fourier plot also show a remarkable congruence with the development of the four phases and the most notable events during the vaccination process. Although the way to obtain the valence of each tweet may look rather coarse, there is a considerable degree of theoretical sophistication in this evaluation of emotional valence. Some of the most accepted theories of emotions rely on two-dimensional spaces where valence becomes one of the fundamental dimensions.[45–49] The six basic emotions due to Paul Eckman[50] are generally maintained, although it is also generally accepted the need to enlarge these basic emotions.[51,52]

Sentiment analysis indeed offers an exciting panorama of emerging tools and paradigms to explain the emergence of social moods and emotional contagion phenomena that are so important in our societies, including the current ‘epidemic of loneliness’.[53,54]

Looking at the limitations of the present approach, we have to consider the existing complementarity between the sentiment analysis technic using lexicons, as herein developed, and the machine learning and deep learning models (supervised and unsupervised).[55] Lexicon-based models are to be preferred where the data sets are small and the available computational resources limited under the condition of slightly lower performance.[56] The supervised models perform fine for the specific domain they have been trained. But this specific training becomes an important limitation for addressing different domains or brand-new topics such as the present COVID-19 pandemic. The unsupervised learning approaches do not hinge on the domain or topic of the training data, overcoming the difficulty of labelled training data collection and creation, although they need an extensive learning process and the subsequent computational resources. The hybrid technique is the combination of both lexicon and deep learning approaches. This combination improves the performance of classification, makes the detection and measurement of sentiment at the concept level and provides high accuracy results.[57]

Conclusions

The new approach developed combines machine learning techniques (sentiment analysis and data mining) with multivariate analysis methods (SNA and text mining). Free software, that is very easy to access and use, has been used to do this. We are currently working on a research project aiming at integrating all these software tools into a Decision Support System, easier to use and interpret the results.

The sentiment analysis approach has proven its validity to evaluate the social mood of citizens in different time scales, registering the different clusters that emerged, gauging public states of mind via the collective valence and detecting the prevalence of the different emotions in the successive phases of the pandemic.

The approach has also shown, albeit rather indirectly, social support for public policies. Overcoming the conceptual limitations around the study of emotions may considerably enrich the perspectives and applications of sentiment analysis and similar kinds of studies, particularly thinking in the emerging mental pathologies—and not only in viral pandemics—around the ‘information society’.

Finally, the combination in formal models of objective and subjective information, in this case about the COVID-19 vaccination process in Spain, will provide a more accurate vision of social reality, which will enable a more effective resolution of problems.

Author statements

Ethical approval

This work did not need to be approved by an ethics committee, as we used public information and messages from the social network Twitter.

Funding

This research is partially funded by the ‘Grupo Decisión Multicriterio Zaragoza’ research group (S35-20R) and the project ‘Participación Ciudadana Cognitiva y Decisiones Públicas. Aplicaciones Sociosanitarias’ (Ref. LMP35_21), both supported by grants from the Regional Government of Aragon and FEDER funds. The funders had no role in the design of the study; in the writing of the article, or in the decision to publish the results.

Competing interests

None declared.

Authors' contributions

A.T. contributed to conceptualisation, methodology, software, data curation, formal analysis, and writing, reviewing and editing the article. A.A. contributed to formal analysis and reviewing and editing the article. J.M.M.-J. contributed to conceptualisation, methodology, reviewing and editing, and funding acquisition. J.N. contributed to conceptualisation, methodology, software, data curation, formal analysis, and writing, reviewing and editing the article.

References

1. Sattar NS, Arifuzzaman S. COVID-19 vaccination awareness and aftermath: public sentiment analysis on twitter data and vaccinated population prediction in the USA. *Appl Sci* 2021 Jun 30;**11**(13):6128.
2. Yousefinaghani S, Dara R, Mubareka S, Papadopoulos A, Sharif S. An analysis of COVID-19 vaccine sentiments and opinions on Twitter. *Int J Infect Dis* 2021 Jul;**108**:256–62.
3. Kapoor KK, Tamilmani K, Rana NP, Patil P, Dwivedi YK, Nerur S. Advances in social media research: past, present and future. *Inf Syst Front* 2018 Jun 1;**20**(3):531–58.
4. Öztürk N, Ayvaz S. Sentiment analysis on Twitter: a text mining approach to the Syrian refugee crisis. *Telematics Inf* 2018 Apr;**35**(1):136–47.
5. https://twitter.com/; 2022.
6. https://www.statista.com/statistics/303681/twitter-users-worldwide/; 2022.
7. https://www.internetlivestats.com/twitter-statistics/; 2022.
8. Chandra Pandey A, Singh Rajpoot D, Saraswat M. Twitter sentiment analysis using hybrid cuckoo search method. *Inf Process Manag* 2017 Jul;**53**(4):764–79.
9. Zimbra D, Abbasi A, Zeng D, Chen H. The state-of-the-art in twitter sentiment analysis. *ACM Trans Manag Inf Syst* 2018 Sep. 5;**9**(2):1–29.
10. Naseem U, Razzak I, Musial K, Imran M. Transformer based deep intelligent contextual embedding for twitter sentiment analysis. *Future Generat Comput Syst* 2020 Dec;**113**:58–69.
11. Mendon S, Dutta P, Behl A, Lessmann S. A hybrid approach of machine learning and lexicons to sentiment analysis: enhanced insights from twitter data of natural disasters. *Inf Syst Front* 2021 Sep. 14;**23**(5):1145–68.
12. Giachanou A, Crestani F. Like it or not. *ACM Comput Surv* 2016 Nov 11;**49**(2):1–41.
13. Antonakaki D, Fragopoulou P, Ioannidis S. A survey of Twitter research: data model, graph structure, sentiment analysis and attacks. *Expert Syst Appl* 2021 Feb;**164**:114006.
14. Hasan MdR, Maliha M, Arifuzzaman M. Sentiment analysis with NLP on twitter data. In: *2019 international conference on computer, communication, chemical, materials and electronic engineering (IC4ME2)*. IEEE; 2019. p. 1–4.
15. Zahra K, Imran M, Ostermann FO. Automatic identification of eyewitness messages on twitter during disasters. *Inf Process Manag* 2020 Jan;**57**(1):102107.
16. Agarwal A, Singh R, Toshniwal D. Geospatial sentiment analysis using twitter data for UK-EU referendum. *J Inf Optim Sci* 2018 Jan 2;**39**(1):303–17.
17. Ilyas SHW, Soomro ZT, Anwar A, Shahzad H, Yaqub U. Analyzing Brexit's impact using sentiment analysis and topic modeling on Twitter discussion. In: *The 21st annual international conference on digital government research*. New York, NY, USA: ACM; 2020. p. 1–6.
18. Somula R, Dinesh Kumar K, Aravindharamanan S, Govinda K. *Twitter sentiment analysis based on US presidential election 2016*. 2020. p. 363–73.
19. Chaudhry HN, Javed Y, Kulsoom F, Mehmood Z, Khan ZI, Shoaib U, et al. Sentiment analysis of before and after elections: twitter data of U.S. Election 2020. *Electronics (Basel)* 2021 Aug 27;**10**(17):2082.
20. Budiharto W, Meiliana M. Prediction and analysis of Indonesia Presidential election from Twitter using sentiment analysis. *J Big Data* 2018;**5**(1):51. C.
21. Sharma A, Ghose U. Sentimental analysis of twitter data with respect to general elections in India. *Proc Comput Sci* 2020;**173**:325–34.
22. Patel R, Passi K. Sentiment analysis on twitter data of world cup soccer tournament using machine learning. *IoT* 2020 Oct 10;**1**(2):218–39.
23. Asif M, Ishtiaq A, Ahmad H, Aljuaid H, Shah J. Sentiment analysis of extremism in social media from textual information. *Telematics Inf* 2020 May;**48**:101345.
24. Kasson E, Singh AK, Huang M, Wu D, Cavazos-Rehg P. Using a mixed methods approach to identify public perception of vaping risks and overall health outcomes on Twitter during the 2019 EVALI outbreak. *Int J Med Inform* 2021 Nov;**155**:104574.
25. Garcia K, Berton L. Topic detection and sentiment analysis in Twitter content related to COVID-19 from Brazil and the USA. *Appl Soft Comput* 2021 Mar;**101**:107057.
26. Manguri K H, Ramadhan R N, Mohammed Amin R, Twitter P. Sentiment analysis on worldwide COVID-19 outbreaks. *Kurd J Appl Res* 2020 May 19:54–65.
27. Turón A, Aguaron J, Escobar MT, Moreno-Jiménez JM. Decision analysis in e-cognocracy using dynamic social networks. In: *ICDSST2020 conference*. Zaragoza; 2020.
28. Scott J. Social network analysis. *Sociology* 1988 Feb 2;**22**(1):109–27.
29. West R, Paskov HS, Leskovec J, Potts C. Exploiting social network structure for person-to-person sentiment analysis. *Trans Assoc Comput Linguist* 2014 Dec;**2**:297–310.
30. https://www.sanidad.gob.es/profesionales/saludPublica/prevPromocion/vacunaciones/covid19/docs/COVID-19_EstrategiaVacunacion.pdf; 2022.
31. https://www.sanidad.gob.es/profesionales/saludPublica/ccayes/alertasActual/nCov/vacunaCovid19.htm; 2022.
32. https://www.vacunacovid.gob.es/; 2022.
33. https://www.openstreetmap.org; 2022.
34. http://gephi.org; 2022.
35. Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. *J Statist Mech Theory Exp [Internet]* 2008;**2008 Oct**(10):P10008. https://doi.org/10.1088/1742-5468/2008/10/p10008.
36. Jacomy M, Venturini T, Heymann S, Bastian M. ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PLoS One* 2014 Jun 10;**9**(6):e98679.
37. Jockers M. *Syuzhet: extracts sentiment and sentiment derived plot arcs from text, R package version 1.0-6*. https://cran.r-project.org/web/packages/syuzhet/syuzhet.pdf; 2020.
38. Bravo-Marquez F, Frank E, Mohammad SM, Pfahringer B. Determining word-emotion associations from tweets by multi-label classification. In: *2016 IEEE/WIC/ACM international conference on Web intelligence (WI)*. IEEE; 2016. p. 536–9.
39. Mohammad SM, Turney PD. Crowdsourcing a word-emotion association lexicon. *Comput Intell* 2013 Aug;**29**(3):436–65.
40. Mohammad SM. Sentiment analysis. In: *Emotion measurement*. Elsevier; 2021. p. 323–79.
41. de las Heras-Pedrosa C, Sánchez-Núñez P, Peláez JI. Sentiment analysis and emotion understanding during the COVID-19 pandemic in Spain and its impact on digital ecosystems. *Int J Environ Res Publ Health* 2020 Jul 31;**17**(15):5542.
42. Liang H, Fung ICH, Tse ZTH, Yin J, Chan CH, Pechta LE, et al. How did Ebola information spread on twitter: broadcasting or viral spreading? *BMC Publ Health* 2019 Dec 25;**19**(1):438.
43. Rufai SR, Bunce C. World leaders' usage of Twitter in response to the COVID-19 pandemic: a content analysis. *J Public Health* 2020 Aug 18;**42**(3):510–6.
44. Wang Y, Croucher SM, Pearson E. National leaders' usage of twitter in response to COVID-19: a sentiment analysis. *Front Commun (Lausanne)* 2021 Sep. 8:6.
45. Handel S. *Classification of emotions*. http://www.theemotionmachine.com/classification-of-emotions/; 2012.
46. SeshathriAathithyan S, Sriram Mv, Prasanna S, Venkatesan R. Affective — hierarchical classification of text — an approach using NLP toolkit. In: *2016 international conference on circuit, power and computing technologies (ICCPCT)*. IEEE; 2016. p. 1–6.
47. Russell JA, Barrett LF. Core affect, prototypical emotional episodes, and other things called emotion: dissecting the elephant. *J Pers Soc Psychol* 1999;**76**(5):805–19.
48. Russell JA. A circumplex model of affect. *J Pers Soc Psychol* 1980;**39**(6):1161–78.
49. Barrett LF. Solving the emotion paradox: categorization and the experience of emotion. *Pers Soc Psychol Rev* 2006 Feb 21;**10**(1):20–46.
50. Ekman P. An argument for basic emotions. *Cognit Emot* 1992 May 7;**6**(3–4):169–200.
51. Ekman P. Basic emotions. In: Dalgleish T, Power M, editors. *Handbook of cognition and emotion*. Sussex (UK): John Wiley & Sons; 1999. p. 45–60.
52. Navarro J, Marijuán PC. The natural, artificial, and social domains of intelligence: a triune approach. *Proc West Mark Ed Assoc Conf* 2022. In press.
53. Bernal AI. Audiencias y temas de noticias en medios online y Twitter. El caso de elpais.com. In: *V congreso internacional de Ciberperiodismo y Web 20*. Audiencias Activas y Periodismo.; 2013.
54. Navarro J, Turón A, Altuzarra A, Lahoz-Beltra R. Comparative sentiment analysis of COVID-19: a machine learning approach. In: *7th ICDSST 2021 (international conference on decision support System technology)*. Loughborough (Reino Unido); 2021.
55. Verma B, Thakur RS. *Sentiment analysis using lexicon and machine learning-based approaches: a survey*. 2018. p. 441–7.
56. Catelli R, Pelosi S, Esposito M. Lexicon-based vs. Bert-based sentiment analysis: a comparative study in Italian. *Electronics (Basel)* 2022 Jan 26;**11**(3):374.
57. D'Andrea A, Ferri F, Grifoni P, Guzzo T. Approaches, tools and applications for sentiment analysis implementation. *Int J Comput Appl* 2015;**125**(3).

International Journal of Disaster Risk Reduction 91 (2023) 103694

Contents lists available at ScienceDirect

International Journal of Disaster Risk Reduction

journal homepage: www.elsevier.com/locate/ijdrr

Press media impact of the Cumbre Vieja volcano activity in the island of La Palma (Canary Islands): A machine learning and sentiment analysis of the news published during the volcanic eruption of 2021

J. Navarro [a], J. Urias Piña [b], F. Magdaleno Mas [d], R. Lahoz-Beltra [b,c,*]

[a] Grupo Decisión Multicriterio Zaragoza (GDMZ), Departamento de Economía Aplicada, Facultad de Economía y Empresa, Universidad de Zaragoza, 50003, Zaragoza, Spain
[b] Department of Biodiversity, Ecology and Evolution (Biomathematics), Faculty of Biological Sciences, Complutense University of Madrid, 28040 Madrid, Spain
[c] Modeling, Data Analysis and Computational Tools for Biology Research Group, Complutense University of Madrid, 28040, Madrid, Spain
[d] Ministry for the Ecological Transition and the Demographic Challenge, And Technical University of Madrid, Spain

ARTICLE INFO

Keywords:
Sentiment analysis
Machine learning
Volcanic eruption management
Disaster journalism analysis

ABSTRACT

In this work we have used as a source of information a large sample of the press articles published during 2021 about the eruption of the Cumbre Vieja volcano in the island of La Palma (Canary Islands). In contraposition, the scientific papers evaluating different facets of natural disasters have preferentially used social networks as a source of information. Herein we have shown how the emotions and sentiments expressed in press media can be efficiently analyzed via AI techniques to better assess the social impact of a disaster at the time it takes place. We have also gauged the usefulness of different classifiers combining sentiment analysis with multivariate statistical analysis and machine learning techniques. By applying this methodology, we were able to classify a newspaper article within a certain time frame of the eruption, and we observed significant differences between local news published in Spanish and those of foreign newspapers written in English. We also found different emotional trajectories of articles by applying the Fourier transform onto the inner "valence" progress along each article narrative time. In addition, there appeared a significant relationship between the surface area occupied by lava and the emotions and sentiments expressed in the articles—many other correlations and causalities could be explored too. The main findings of this research may constitute a helpful resource for a better understanding of the way press media react to volcanic activity, and may guide in public decision-making under different temporal horizons, including the design of improved strategies in the risk reduction domain.

1. Introduction

Cumbre Vieja volcano, the most active of the Canary Islands according to existing historical records, came into eruption on

* Corresponding author. Department of Biodiversity, Ecology and Evolution (Biomathematics), Faculty of Biological Sciences, Complutense University of Madrid, 28040 Madrid, Spain.
E-mail address: lahozraf@ucm.es (R. Lahoz-Beltra).

https://doi.org/10.1016/j.ijdrr.2023.103694
Received 16 November 2022; Received in revised form 23 March 2023; Accepted 12 April 2023
Available online 19 April 2023
2212-4209/

September 19, 2021 (Fig. 1), following five decades of repose [1]. But Cumbre Vieja is not properly a volcano—it is a monogenetic volcanic field. Actually, in the geological records the erupting volcano has been called “Tajogaite”, and it is a monogenetic volcano located inside the Cumbre Vieja volcanic field. But herein we maintain the popular name that has been used in press articles. Before the eruption, the Spanish Geographic Institute (IGN) and the Canarian Vulcanologic Institute (IVC) had detected from September 11, 2021, an intensified seismicity in the area, with thousands of seismic episodes of different magnitude, which began at a depth of over 20 km and progressively ascended towards the surface. The volcano eruption lasted 85 days and 8 h, originating damages estimated by the regional government of circa 900 million €. During that period, 2988 buildings were destroyed, from which 1345 had residential use, and 7000 people had to leave their houses [2]. Satellite information from EU Copernicus programme [3] showed that 1219 ha of land were covered by the volcanic lava, from which 370 corresponded to cropping areas—bananas mainly, also vineyards and avocado groves.

Cumbre Vieja had been studied for decades by different authors regarding its unrest signals [4], helium emission and concentration in its soils [5], its dynamics of diffuse carbon dioxide emissions [6], and the potential effects of such reactivation, such as near- and far-field tsunamis [7–9]. The interest of scientists and other segments of society for the governing processes and the predictable aftermaths of the volcano eruptions has kept going for many years, in the light of the aforementioned works and many others. Scientific and social awareness and attraction by Cumbre Vieja, and by volcanoes in general, is well-known. In fact, the term “volcanologist paradox” [1] was coined to explain the interpretive tension arising by the coexistence of specialists and the laymen fascination for eruptions, which also occurs in the evaluation of their consequences—in the human and monetary costs associated. Minimizing these conceptual tensions would imply working in the construction of a reinforced trust and collective involvement, particularly among expert teams and local people, including the press media too.

Social grieving for the damages created by the Cumbre Vieja eruption has been widely shown by local, national, and international media, and some authors have analyzed the relevance of adding new dimensions (mostly related to empathy and support) beyond the current vision of grief, by incorporating more adaptive and resilient processes within the personal and social sentiment of coping [10]. In this line, understanding natural events and their interaction with people’s emotions and ways of life would oblige researchers and managers to study and interpret different environments: the closer physical environments of those more affected by the eruption, the affective environment (family, friends, neighborhood, school, etc.), and the communication environment as well. It would contribute to disseminate a more solid knowledge of the natural and human processes along the volcanic eruption [11]. In this respect, by analyzing how different media follow the natural dynamics and their immediate consequences, deeper insights for accurate communication of public policies and risk-taking decisions could be achieved.

1.1. On the coverage of different media

A thorough and systematic approach to the media impact of this volcanic eruption, and of other volcanoes, looks feasible and

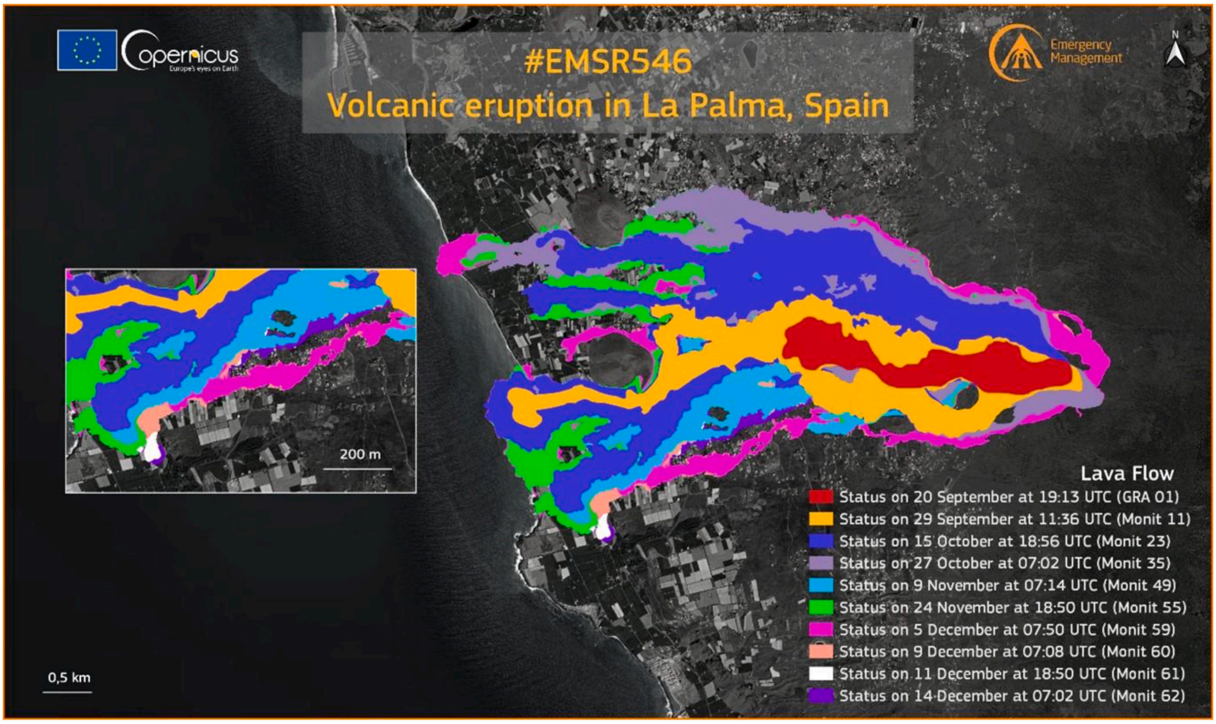

Fig. 1. Eruption of the Cumbre Vieja (“Tajogaite”) volcano on the island of La Palma (Canary Islands) during 2021. Evolution of the lava flow from September 20 to December 14 (Source: Copernicus. https://www.copernicus.eu/en/news/news/observer-copernicus-eyes-la-palma-eruption).

helpful via the new sentiment analysis tools. In the construct of the social reality around an eruption in a complex semi-urban population, like in the Cumbre Vieja eruption, there were many participant voices – speeches of volcanologists, inhabitants of affected areas, vulnerable elders, and children, affected farmers, tourists, different administrations, etc. – which were largely reflected and reconfigured within the coverage of the different media. In this regard, the main difference between press media approaches (newspapers & news media) and social media approaches (social networks) is that there is a surplus of reflection and analytical thought inside the *curated* contributions of journalists, notwithstanding the ever-present temptation to sensationalize, in comparison with the immediacy and emotional spontaneity of social media contents.

So, thinking in terms of decision-making and preparedness for future risk reduction, the information distilled from press media would be, in Daniel Khaneman's terms of System 1 immediacy (S1) and System 2 reflexivity (S2) [12], halfway positioned in between these two systemic extremes. Thus, the S1 cascades of tweets or similar platforms and the S2 long-term commissioned reports, scientific articles, and meta-reviews, with the press media situated in between, would constitute an information continuum of social communication modalities. In this work we highlight potential advantages of press media for public decision bodies and risk management—including the necessity of getting rid of the vast misinformation campaigns that plague social media and social networks: fake news, trolls, bots, etc. This is an important concern when handling masses of tweets or from other platforms.

In the above sense, deceptive social bots—automated or semi-automated accounts designed to impersonate humans—have been effectively exploited for these kinds of abuse [13,14]. The spread of fake news and infodemic have increased dramatically worldwide in the last years, especially during the COVID-19 pandemic [15]. Such unverified and inaccurate information is generally referred as encompassing misinformation (false information created without any harmful intention), disinformation (false information deliberately created to harm an entity), or malinformation (information based on reality, created to inflict harm on an entity) [16]. Very often, during a disaster the confusion created leads to the publication of misinformation in social media [17]. During the Cumbre Vieja eruption, the volcanologist Dr Janine Krippner asked on Twitter to avoid the dissemination of tweets which respond to rumors related with the formation of tsunamis. In addition, the information published by social media users is subject to the ups and downs of events and mood changes in the affected community [18]. Thus, the fluctuations undergone by the information disseminated by social media can be useful in some cases, e.g., during Hurricane Dorian [18], but not for others. Furthermore, during the eruption of the Hunga Tonga-Hunga Ha'apai submarine volcano, the submarine cables were severely damaged cutting off Tonga (a Polynesian archipelago) communications with the outside world [19].

1.2. Social reactions to volcanic activity

Social perspectives about volcanoes in our times have been largely influenced by the way they have been portrayed in literature, films, and media [20]. It is also worth recalling that many cultures and civilizations which coexisted with volcanoes came into contact with them in pre-scientific ages, which contributed to the development of particular ways to explain their dynamics, usually involving mythological approaches that reflected fearful and even apocalyptic feelings towards eruptions [21]. Potentially, some ancient texts are susceptible of comparative analysis following our approach, particularly when different narratives have survived. In our times, the public reactions to volcanic activity are mostly based on its intensity and variability, on the risk of collateral destructive events (fires, earthquakes, tsunamis), and in the preparedness and potential vulnerability of neighboring populations [22]. The press information could efficiently contribute to bring these factors on the table and to discuss them more meaningfully. As argued, it can straightforwardly sidestep and minimize the risks of misinformation via fake news and media bots.

About the emotions and feelings displayed in volcanic eruptions, the perception of danger is essential. It generates the most powerful emotional reactions [23–25], gauged by the proximity to the danger area and other risk factors. These human emotional reactions have strong evolutionary roots [26]. They can be classified in fixed action patterns, flexible perception-action patterns, and superstructural perception-action patterns [27]. Depending on the level of danger, reactions may rapidly escalate towards the most basic or "fixed" emotions. In general, negative sentiments belong to the superstructural type, related to a combination of different personal, social and physical aspects, biased by income, demographic layers, livelihood type, level of education, culture, religion and perception, among other factors, which also include local and national governance responses. At any rate, the interactions between people and volcanic effects become multifaceted and complex, also comprising positive emotions about the associated risk hazards, which could suggest the necessity of applying an open-risk and multidimensional comprehension about such interactions [28–30], which additionally co-evolve following the natural patterns of perturbation.

Evaluating and monitoring sentiments about volcanic activity may thus be relevant to better understand such co-evolution and better prepare society for future volcanic risks. Sentiment analyses or opinion mining of different events and processes, including volcanic activity, have been increasingly explored [31]. Current opinion mining approaches are manifold, focused on the classification of positive/negative valences (polarity sentiment analysis [32]) and on the typification of the different levels of detail which may be applied to the analysis. By introducing more specifically oriented statistical methods, some common challenges in opinion mining may be overcome: (i) the conflict induced by the fact that an opinion word could be deemed as positive or negative according to specific situations; (ii) the probability that people could not transmit their opinions in the same way under different circumstances; (iii) the relative absence of sentiment analyses in languages different from English or Chinese [33], which are the dominant languages, up to this date, in these studies.

In essence, we are trying to answer the following question: Do the emotions and sentiments expressed by a journalist in a newspaper article reflect the date or period in which the article was written and chart the devastation effects caused by a catastrophe or natural disaster? We respond positively and provide robust analytical tools. By counting with the possibility of classifying press news within a certain episode or period of time during the development or evolution of a natural disaster, with the emerging differences between national and international clustering effects, and with the corresponding valence scrutiny via Fourier transform, new avenues may be open for the careful analysis of social impact, for better assessing the changing of social mood, and for the design of improved risk-reduction communication protocols. Anecdotally, the social impacts of an environmental disaster or natural catastrophe of the past might also be gauged many years after the catastrophe, based on the different narratives preserved.

2. Review of literature

Volcanic eruptions have been the subject of a number of studies. Each episode has generated specific approaches according to its own geophysical characteristics. For instance, the Fagradalsfjall eruption, in Iceland 2021, particularly affected atmospheric dynamics [34]; the Hunga Tonga-Hunga Ha'apai unleashed an enormous submarine eruption accompanied with important tsunamigenesis [35, 36]; the Anak Krakatu eruption in Indonesia 2018 generated a devastating tsunami [37]. Similarly, recent studies during the Tajogaite (Cumbre Vieja) eruption have covered geological/geophysical aspects [38], as well as the general impact of the eruption, its aftermath, and the long-term challenges [39].

Regarding sentiment analysis, it has been applied to the study of numerous crises and natural disasters, particularly during the pandemic caused by COVID-19, taking as the usual source of information micro-blogging services, i.e., Twitter [40–42]. In these cases, social networks allow a faster dissemination of the posted information, highlighting damages and help needed "in real time" [18]. The pros and cons of Twitter based approaches become evident. Subsequently, sentiment analysis techniques applied to social networks have been a useful tool in many quantitative analyses of crises and disasters [43]. See for instance text analysis for volcano monitoring [44], disaster response and recovery [45], Syrian chemical attack [17], COVID-19 pandemic [46,47], L'Aquila's earthquake [48], hazard crises responses [49], risk detection through crisis information [50], and hurricane Dorian social impact [18].

3. Methodology

The methodological approach was based on the combination of sentiment analysis with machine learning techniques and multi-variate statistical models. The general protocol was similar to the approach used in previous studies in which these authors analyzed the conversations of a chat bot with other bots or with a human interlocutor [51] and about the effects of the COVID-19 pandemic [47, 52]. The chart in Fig. 2 shows the methodological procedure, which basically consisted of three steps. First, news were collected from press media sources, and their texts were normalized (depurated and cleaned); then, they were object of sentiment analysis; and finally, the results obtained were processed by means of multivariate statistical analysis and machine learning methods.

3.1. News collection and depuration

A total of 158 press items were collected and analyzed. The press articles were classified into two groups, one consisting of press articles written in Spanish ($n = 87$) and the other group of articles written in English ($n = 71$). The press articles were collected from the beginning of the volcano eruption (September 19, 2021) until the official date of the end of the eruption (December 13, 2021), with an extra period of about two weeks in order to compile a sufficient number suitable for statistical analysis. Thus, the articles were

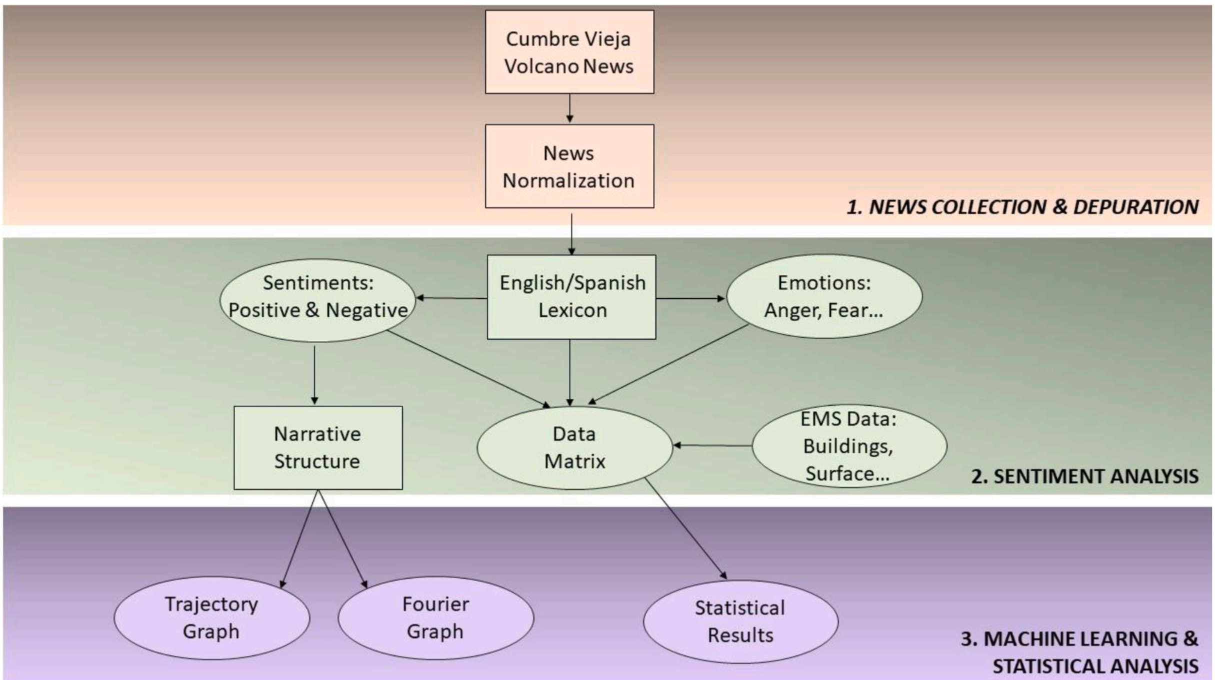

Fig. 2. Methodological flow diagram.

collected until December 26 and December 31 for the news written in Spanish and English, respectively. The final choice of 158 press articles, 87 articles published in the Spanish press and 71 published in foreign media, is statistically appropriate because sample sizes are both representative ($n > 30$) and close to each other, being a balanced design.

The sources of the articles were official web pages of reliable press media, newspapers, TV, and radio media. The Spanish and foreign press media were chosen on the basis of whether they were published in both print and online, as well as on their relevance. A similar criterion, relevance, was followed to choose radio and TV channels. Otherwise, sports newspapers, e.g. Marca, were selected because they reported the events in the local 'La Palma' on time, having more diffusion than the general press. The press articles in Spanish were extracted from the following information sources: As, Diario de Navarra, El Pais, La Vanguardia, El Mundo, RTVE, Telecinco, Cadena Ser, La Razon, Antena3, Marca, and El Confidencial. The sources from which we obtained the press releases published in English were: Reuters, Garda, BBC UK, BBC, Aljazeera, NY Times, The Guardian, CNN Reuters, The Local, Euronews, Euroweekly News, NBC Connecticut, New Indian Express, ITV News, Voanews, Volcano.si.edu, Surinenglish, UK Yahoo News, Accuweather, Earthsky, VolcanoDiscovery, Globalnews, CNN, Dailymail UK, Phys.org, CBS News, Earth observator. In this mix, which we think is also representative internationally of widely followed media, there is a small but significant presence of specialized sources (it may contribute to explain some of the risk appreciation differences found).

All the collected news were 'depurated'. Since their texts contained words in different tenses, plurals, or derived from other words, they were normalized transforming the words into their basic forms. Next, the texts were 'cleaned up', a process that included different tasks such as orthographic correction, elimination of punctuation marks and special characters, conversion of acronyms to regular expressions, conversion of capital letters to lower case, etc.

3.2. Sentiment analysis

The sentiment analysis was performed separately in two groups of press articles depending on the language used. The press articles were analyzed in different steps applying text mining by means of the Syuzhet 1.0.6 package [53] and RStudio 1.1.419, as shown in Fig. 2.

In both cases (English and Spanish), the procedure applied was similar. Once the text of a press article was normalized (depurated and cleaned), it was fragmented into smaller strings or sentences. The result of this preliminary analysis was a sentiment vector whose length, i.e., the number of elements, was the number of paragraphs that frame the text from each press article. Thereafter, each element of the vector was given the value that corresponded to the evaluation of the emotion and sentiment words in the analyzed item. The sentiment valence was calculated for each one of the sentences obtaining its value as the difference between the number of positive and negative words. In addition, it was also obtained the number of words associated with the basic emotions, calculating the percentage of words associated with each emotion expressed in the text. This sentiment analysis procedure was performed separately in the two groups of press articles depending on the language used.

The sentiment analysis was conducted with NRC Word-Emotion Association Lexicon Version 0.92 [54,55]. This lexicon is a list of English/Spanish words and their associations with eight basic emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust) and two sentiments (negative and positive). In total, the lexicon contains 14,182 unigrams (words) and more than 25,000 senses. The NRC lexicon can be explored deeply through an interactive visualization on the NRC website [56], where one can look for the number of words associated with each emotion, word-sentiments associations, and word-emotions associations. The Spanish version of the lexicon has already been used to assess mood evolution in the COVID-19 vaccination process in Spain [47].

Next, from the values of the sentiment vector– the valences of the sentences– we examined how emotions were distributed in the text, obtaining an overall statistical assessment of each press article. To this end, several univariate sample statistics were obtained: the minimum value (Min), the first quartile (Q_1), the median (Me), the mean ($\overline{x}$), the third quartile (Q_3), and the maximum value (Max). Another feature analyzed in the articles was the story narrated in the text and how the frequency of words expressing positive or negative sentiments changed over the course of the article. The result of this analysis was a trajectory graph—a plot of the variation of the emotional valence with respect to the narrative time.

Afterwards, to eliminate the extreme values of sentiments we applied the Fourier transform, converting the trajectory graph into another equivalent graph independent of the length of a press article [57]. Based on this graph, it was possible to find out which sentences of a particular article item expressed a positive or negative emotion, whether the narrative of a specific story evolved towards a happy or a sad ending, etc. In previous studies [51] these authors found that the graph of the Fourier transform, plotting the emotional variation of valence with respect to the narrative time, followed one of four possible elementary patterns (Positive, Negative, O_Negative, O_Positive), which are the Fourier terms describing the elementary 'arc' evolution of a narrative. Lastly, once the Fourier plot of each article was obtained, we analyzed with a chi-square test of independence whether there was a relationship between the number of Fourier plots of each class (Positive, Negative, O_Negative, O_Positive) and the month during the volcanic eruption in which the article was published (September, October, November, December).

3.3. Statistical and machine learning analysis

3.3.1. Statistical analysis

As a result of the sentiment analysis, a data matrix was obtained for the total of analyzed press articles. In this matrix, there were 16 columns or predictor variables whose values were six univariate sample statistics of the sentiment vector (Min, Q_1, Me, $\overline{x}$, Q_3, Max) as well as the total number of words associated with eight emotions (anger, fear, anticipation, trust, surprise, sadness, joy and disgust) and two sentiments (negative and positive). In this matrix, the units of analysis – the press articles – were placed in rows.

Once completed the previous stage, the data matrix was analyzed by applying multivariate statistical analysis and machine learning methods. The aim of these analyses was to answer a series of relevant questions. In a press article about a natural phenomenon that represents a disaster for the involved population, such as the volcanic eruption of Cumbre Vieja and based on the analysis of the sentiments expressed in the article, is it possible to estimate the time period of the eruption at which the article was published? Is there any difference in the sentiments expressed in a press article depending on whether the journalist or the news media are local or foreign? In a newspaper story, is there any stochastic relationship between the predictor variables resulting from the sentiment analysis and the variables related with the eruption effects, such as the surface area occupied by the lava?

In order to answer these questions, we performed a principal component analysis [58]. The goal was to simplify the information provided by the 16 predictor variables in the data matrix, so to choose a smaller number of predictors and to understand the data matrix in a simpler way, reducing the dimension space. For this purpose we obtained the eigenvalues—the proportion of variance explained by the new predictor variables or principal components.

Next, our goal was to study the suitability of differente classifiers in order to have a predictive model capable of classifying the press news in one of the four months in which the Cumbre Vieja volcanic eruption occurred. Classifiers are procedures or algorithms that classify the news in different classes or groups, in particular the period of time in which the article was published. If the classification of a given article was correct, then the classifier would have accurately detected the period in which the news item was published via the sentiments expressed in its text.

3.3.2. *Combining statistical and machine learning models*

In our study we chose the following classifiers for evaluation: discriminant analysis, perceptron neural network, and logistic regression classifiers; all of them are very common techniques in multivariate statistical analysis and machine learning. Moreover, at present these techniques are included in most of the current statistical packages.

- *Discriminant analysis*. It was conducted by constructing Fisher's linear discriminant functions [59] which maximize the separation of groups. The goal was to find two or more linear combinations of predictor variables, among the 16 variables resulting from the sentiment analysis, with which we could classify a press article in a particular group representing the months in which the Cumbre Vieja volcano eruption took place.
- *Perceptron neural networks*. Artificial neural networks are connectionist machine learning models inspired by the neural circuits of the brain [60] allowing the classification of objects. They consist of nodes representing neurons arranged in layers, with the networks usually comprising an input layer and an output layer, and including one or more intermediate or hidden layers. In the present work we have used a neural network known as a multilayer perceptron network (MLP). In a MLP neural network, the input layer would receive for a certain press item the values of the 16 prescriptor variables, it would internally process that information, and the output layer would obtain the network's response. The output of the network was in this case the month in which the notice about the volcanic eruption was published. Based on a subset of press articles from the data matrix, the MLP network was previously trained to recognize and classify news correctly by applying an algorithm known as backpropagation method [60]. By means of this algorithm, the weights of the connections between nodes were modified, following what is known as supervised learning. The topology of the network included an input layer with 16 neurons, a hidden layer formed by 2 or 3 neurons, and an output layer of 4 neurons corresponding to the months of September, October, November, and December. The 16 neurons in the input layer accounted for the 16 variables (anger, anticipation, disgust, fear, joy, sadness, surprise, trust, negative, positive, Min, Q_1, Me, $\bar{x}$, Q_3 and Max), thus with the number of words associated with the eight emotions, two sentiments, and the sample univariate statistics of the sentiment vector. The activation functions were the hyperbolic and sigmoid tangent for the hidden and output layers, respectively. The gradient slope algorithm was used with an initial learning rate equal to 0.4, and learning was evaluated with the sum of squares as an error function. We compared the appropriateness of the discriminant analysis and the perceptron neural network classifying a press article in a given period of time in two different scales. On the one hand, the articles were classified into one of the four months (September, October, November or December) in which the volcanic eruption took place. On the other hand, we decided to study the suitability of the classifiers sorting the press articles into two groups or clusters. One group '0' included the press articles published in the months of September and October, during the initial stage of the eruption. The other group '1' represented a final stage defined by the articles published during the months of November and December.
- *Logistic regression model*. Since in the latter case the output of the classifier was either 0 or 1, we included the study of other classifier, using the logistic regression model [61] as one of the procedures to classify press releases into each group. With this type of regression model it was possible to predict the outcome of a categorical variable, in our case 0 for the initial stage of the eruption (September, October) or 1 for the final stage (November or December), as a function of the independent or predictor variables.
- *Probabilistic neural network*. Further we studied the suitability of another classifier, a probabilistic neural network, in order to classify each news into one or another group according to the language in which it was written. We classified the articles by means of a probabilistic neural network classifier [62] because we found sentiment differences between the press articles depending on the language in which they were written. The procedure used to carry out the present non-parametric classification method [63] involved the estimation of a density function for each group of articles. The estimation was constructed using a Parzen window, a procedure that weights the observations of each group of articles according to the distance from their location. In this study the network comprised four layers of neurons, i.e. input, pattern, summation, and output layers, which were composed of 2, 10, 2, and 2 neurons respectively. The 2 neurons of the input layer were standing for the predictor variables, so in the present study only the positive and negative sentiments. The 10 neurons of the pattern layer calculated the contribution of the input variables to the density function of each news group. At last, the 2 neurons of the sum layer assigned each press article to one or the other news group, a decision which was obtained and displayed in the 2 neurons of the output layer.

Finally, since the language of the source of information affected the sentiment analysis outcomes, we evaluated the extent to which the emotions and sentiments expressed in the articles were different depending on the language in which they were written. We compared the medians of each emotion between the news written in English and those written in Spanish by means of a Mann-Whitney (Wilcoxon) test. The Wilcoxon test was also applied to compare the medians of positive and negative sentiments in the two groups of press articles.

3.3.3. Statistical relationship between volcanic eruption and press emotions and sentiments

Using Copernicus Emergency Management Service (EMS) as a source of information, we could systematically obtain data on the activity of the volcano during the months of the eruption. For each date, we compiled in a table the surface area (hectares) covered by lava, the number of earthquakes, and the number of buildings destroyed per day.

We studied whether there was a statistical relationship between the surface area covered by lava with the emotions and sentiments expressed in a press article. For this purpose, and due to the presence of multicollinearity among the predictor variables, we applied the multiple linear regression chain method. In this study, however, we only analyzed the surface area covered by the lava to illustrate the magnitude of the natural disaster; the number of earthquakes and the number of buildings destroyed per day could be subject to future research.

In all the above-mentioned works, the principal component analysis, discriminant analysis as well as the logistic regression model, the probabilistic neural network, and the multiple linear regression, their respective models were built by means of STATGRAPHICS Centurion 18 version 18.1.12. Using this statistical package, other statistical tests were also conducted, such as the Wilcoxon test and the multiple linear regression. In addition, the perceptron neural network model was performed with the IBM SPSS Statistics version 22 statistical package.

4. Results

In this Section we processed the data obtained from sentiment analysis. As we described in Fig. 2, there was a narrative structure, a series of variables related to the physical and urban impacts, and a data matrix summarizing all these variables together with sentiment analysis results. Obtaining the eigenvalues of the data matrix (via principal component analysis) appeared as the first step. It would determine whether further predictions based on the data obtained were feasible or not. We had to do that for the news in each language, first in English and then in Spanish.

4.1. Analysis of the news written in English

4.1.1. Principal component analysis: space dimension reduction

In this study we first performed a factorability test by obtaining the Kaiser-Meyer-Olkin measure (KMO) for ideal sampling, assessing whether it was worthwhile to extract reducing factors from a set of variables. Since KMO was equal to 0.835828 (must be at least 0.6) we concluded that in the present study a factorization was feasible. In consequence, a principal component analysis was conducted considering the classification of the press articles during the four months of the eruption. We extracted three principal components, since their eigenvalues (7.6324, 1.9262, 1.5446) were equal or higher than unity. Together the three principal components explained 69.396% of the variability in the original data, with the two principal components PC1 and PC2 (see Table 1) aptly summarizing most variables of the sentiment analysis in each article. The first principal component (PC1) collected the effect of the eight emotions and the two sentiments, their weights being very similar and of positive sign. The second principal component (PC2) aptly captured the effect of the sample statistics calculated from the sentiment vector. Finally, the third principal component (PC3) reflected the rank of the *valence* in a press article: Median, Max, Q1, and Min values of valence.

In Fig. 3 we show the biplot for these two principal components. It is interesting to note a correlation between positive emotions (*joy, trust, surprise, and anticipation*) and positive sentiment (*positive*), as well as between negative emotions (*fear, anger, disgust, and*

Table 1
Table of component weights (English press articles).

	PC1	PC2	PC3
	1	2	3
Anger	0.30906	−0.100829	−0.0888711
Anticipation	0.327952	0.0778098	0.0426475
Disgust	0.282448	−0.0405535	0.00487426
Fear	0.339326	−0.0776851	−0.0609463
Joy	0.303533	0.197328	0.072347
Sadness	0.300654	−0.0600962	−0.105506
Surprise	0.323247	0.0489104	−0.00663019
Trust	0.286234	0.109637	0.11191
Negative	0.330943	−0.146225	−0.0889564
Positive	0.338591	0.0822905	0.0215608
Min	−0.0569136	0.310579	0.309858
Q1	−0.0121977	0.250056	−0.566931
Me	−0.028023	0.418606	−0.524037
Mean	0.00747464	0.464483	0.266309
Q3	−0.0334614	0.478347	−0.127658
Max	0.0595891	0.33497	0.412284

sadness) and negative sentiment (*negative*).

4.1.2. Discriminant analysis classification of press articles

Fig. 4 shows how the impact of the volcanic eruption was reflected in the emotions and sentiments expressed in the articles published all through the months of September, October, November and December. The analysis of emotions combined with the discriminant analysis allowed us to classify each news in the month in which it was written, mirroring the course of the volcanic eruption.

Table 2 shows how the sentiments reflected in the news varied significantly during the four months of the eruption. See also Fig. 4. This was confirmed by a discriminant function statistically significant (*p-value* = 0.0005) that accounted for 76.64% of the variance. A high value of the canonical correlation (0.7295) indicated a strong relationship between the group membership (month in which a press article was published) and the discriminant function values. This was also specified by the eigenvalue ($\lambda = 1.1358$) with a proportion of 23.36% of the total variance (Wilk's Lambda) not explained by differences among months. In consequence, the discriminant function was taking different values in the different months, successfully discriminating among the articles published in each month. The classification table (Table 3) shows how the foreign press articles were correctly predicted in the classification, with an efficacy of 69.01% (weighted arithmetic mean).

Nevertheless, when the press items were only classified into two groups, the initial stage (September and October as group 0) and the final stage (November and December as group 1), then the success rate in the classification was increased. In this case the discriminant analysis correctly classified 76.06% of the articles in one of the two groups (0/1).

4.1.3. Multilayer perceptron network (MLP) classification of press articles

The perceptron neural network (Fig. 5) successfully classified 80% and 78% of the press articles into two groups (0/1) during the training and trying out stages, respectively. We could conclude that both classifiers, discriminant analysis, and perceptron had very similar performance in classifying an article into one of the two groups. It is important to note that the most important predictor variables for the perceptron (Fig. 6) were Q_3 (100%), that is, the third quartile of the valence, plus three emotions: trust (92%), joy (84%), and fear (80%). It was also found that the mean value of valence and the negative sentiments exhibited both equal relevance (74%). In conclusion, the press articles published in English were successfully classified by the perceptron into two groups 0/1, being the most relevant predictor variables fear and the negative sentiments expressed. However, at the same time, the articles written in English reflected 'confidence' (trust) and 'enjoyment' (joy) in the face of a spectacular natural phenomenon.

4.1.4. Logistic regression classification of press articles

In the logistic regression model, the press articles were again classified into the same two groups, obtaining in the chi-square goodness-of-fit test a *p*-value equal to 0.7929. Consequently, a relationship between the group or cluster 0/1 and the independent predictor variables was established by the following equation of the fitted model:

$$G_{English} = \frac{e^{x}}{1 + e^{x}}$$

with x being equal to:

$$x = -0.3326 + 0.0001\ disgust + 0.0004\ anger + 0.0006\ anticipation + 0.0007 fear$$
$$-0.0042\ joy - 0.0017 sadness - 0.0031\ trust + 0.0007\ negative + 0.0010\ positive$$

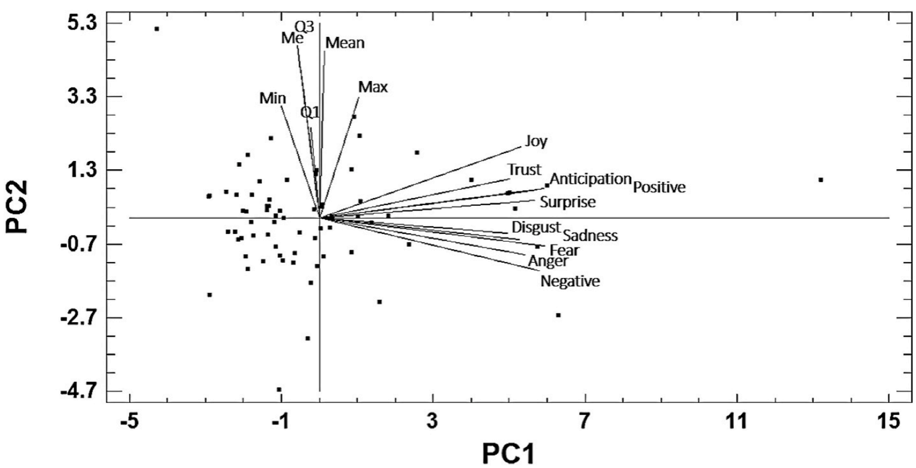

Fig. 3. Biplot of articles written in English showing press articles as a function of the first (PC1) and second (PC2) principal components. The figure shows emotions (joy, trust, anticipation, surprise, disgust, sadness, fear, anger), two sentiments (positive, negative), and sample statistics (Min, Q_1, Me, $\overline{x}$, Q_3 and Max).

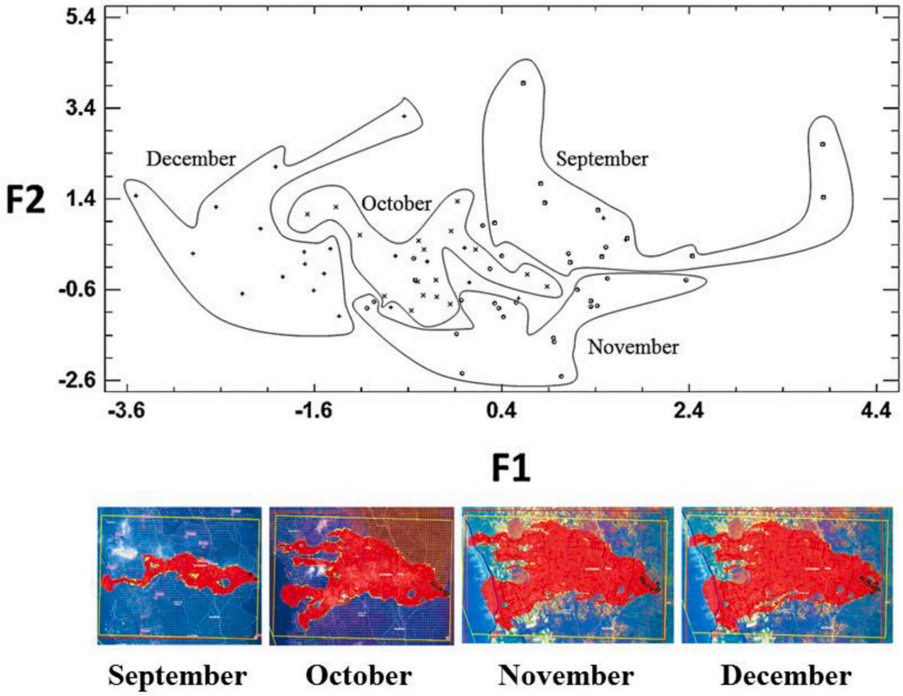

Fig. 4. Monthly classification of foreign press articles (written in English) based on discriminant analysis (F1 and F2 are classification functions). The bottom section of the figure shows the surface occupied by lava during the months of volcanic eruption.

Table 2
Discriminant analysis of English press articles.

Discriminant function	Eigenvalue	Relative percentage	Canonical correlation
1	1.13584	57.73	0.72925
2	0.430197	21.87	0.54845
3	0.40131	20.40	0.53515

Functions	Wilks' Lambda	Chi-squared	d.f.	p-value
1	0.233615	87.2448	48	0.0005
2	0.498965	41.7132	30	0.0757
3	0.713618	20.2445	14	0.1226

Table 3
Classification table of English press articles.

Actual month	Size	*Predicted month*			
		September	**October**	**November**	**December**
September	13	10 (76.92%)	2 (15.38%)	1 (7.69%)	0 (0.00%)
October	17	0 (0.00%)	11 (64.71%)	5 (29.41%)	1 (5.88%)
November	23	3 (13.04%)	3 (13.04%)	15 (65.22%)	2 (8.70%)
December	18	1 (5.56%)	2 (11.11%)	2 (11.11%)	13 (72.22%)

4.2. Analysis of the news written in Spanish

4.2.1. Principal component analysis: space dimension reduction

The principal component analysis with the KMO measure was equal to 0.8607, explaining a 75.46% of the variability. We also obtained the first two principal components, PC1 and PC2 (Fig. 7), showing that their modeling content was similar to the principal component analysis of the English-language news. However, we observed differences between the biplot obtained for the news written in English (Fig. 3) and the biplot for the news published in Spanish (Fig. 7). For the latter, the correlation among the univariate sample

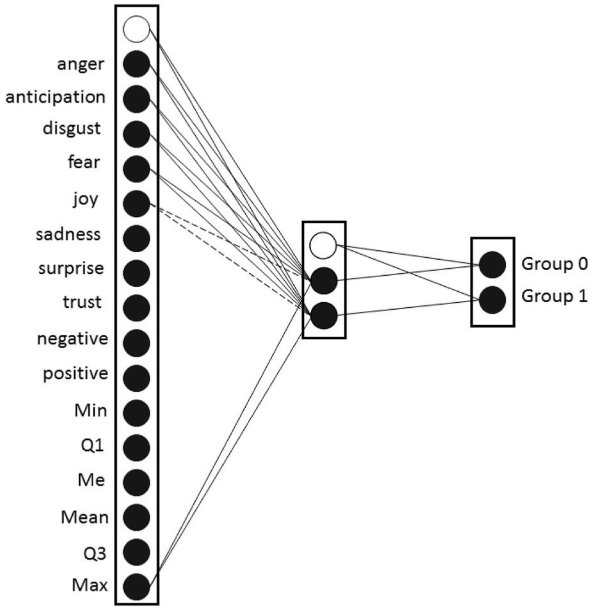

Fig. 5. Multilayer Perceptron Network (MLP) for the classification of news written in English. MLP depicts in the input layer the 16 neurons that receive the values of the prescriptor variables collected in the sentiment analysis step, plus an intermediate or hidden layer formed by 2 neurons, and the output layer with 2 neurons whose activation results in the classification of a news article in Group 0 (articles published in September and October) or in Group 1 (November and December). In the figure the white node represents the voltage or bias used to enhance MLP training via backpropagation, and the black nodes stand for the neurons.

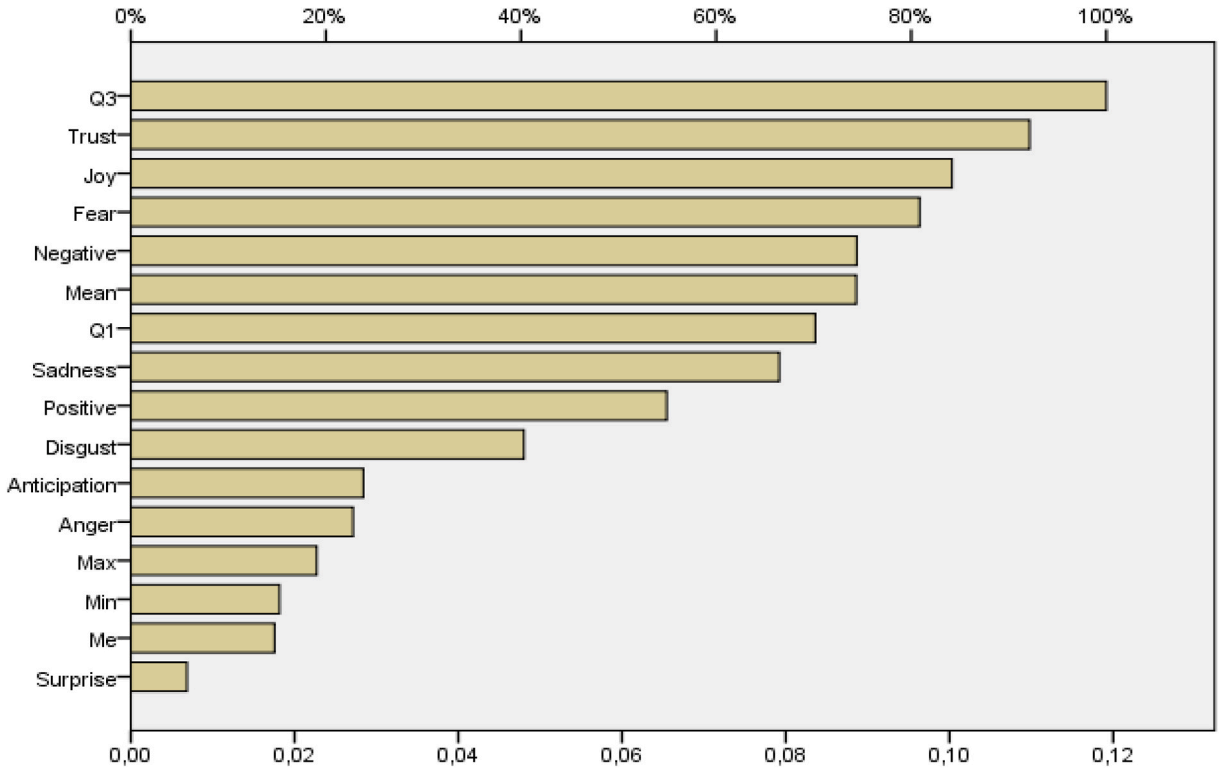

Fig. 6. Bar chart showing the relevance of the predictor variables in the perceptron neural network when classifying the news written in English into two Groups (0 and 1).

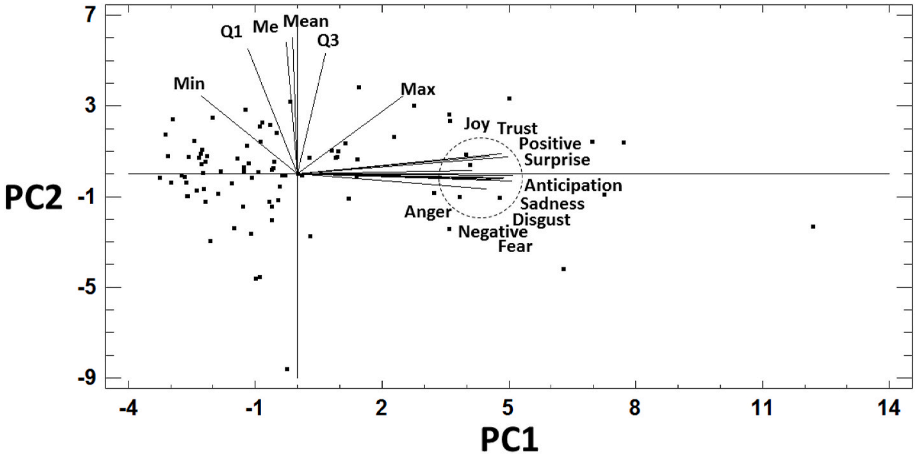

Fig. 7. Biplot of articles written in Spanish showing press articles as a function of the first (PC1) and second (PC2) principal components. The figure shows emotions (joy, trust, anticipation, surprise, disgust, sadness, fear, anger), two sentiments (positive, negative) and sample statistics (Min, Q_1, Me, $\bar{x}$, Q_3 and Max). The circle shows the area in which there is a correlation between emotions (anger, fear, anticipation, confidence, surprise, sadness, joy and disgust) and sentiments (negative and positive).

statistics (Min, Q_1, Me, $\bar{x}$, Q_3, Max) of the sentiment vector was lower than for articles written in English. Also, we found that for Spanish-language articles the correlation between emotions (anger, fear, anticipation, confidence, surprise, sadness, joy and disgust) and sentiments (negative and positive) was higher than for articles written in English.

4.2.2. Discriminant analysis and multilayer perceptron network (MLP) classification of press articles

One of the most surprising outcomes of this work relates to the discriminant analysis results. When we considered the classification of the press articles in Spanish into four groups – the months elapsed from September through December – then the percentage of correctly classified press articles releases dropped to 59.77%.

Similarly, the perceptron neural network (Fig. 8) was less successful in classifying the press articles into two groups or clusters, group 0 (September and October) and group 1 (November and December). In that case, the perceptron classified correctly 59.0% and 69.0% of the articles during the training and try out stages, respectively. However, the discriminant analysis (p-value $= 0.0200$) correctly classified 74.71% of the articles (Table 4). Thus, when we classified the articles into two groups (0/1) the discriminant

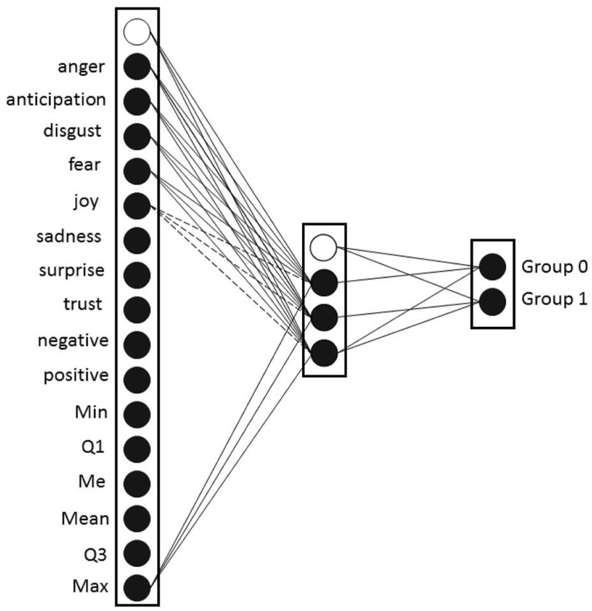

Fig. 8. Multilayer perceptron network (MLP) for the classification of news written in Spanish. MLP depicts in the input layer the 16 neurons that receive the values of the prescriptor variables collected in the sentiment analysis, an intermediate or hidden layer formed by 3 neurons, and the output layer with 2 neurons whose activation results in the classification of a news article in Group 0 (articles published in September and October) or in Group 1 (November and December). In the figure the white nodes represent the voltage or bias used to enhance MLP training via backpropagation, and the black nodes stand for the neurons.

analysis was more efficient than the perceptron neural network. But the percentage of Spanish articles classified correctly was lower than for the foreign press.

In contrast to the articles written in English, another interesting result was that the most important predictor variables in the perceptron classification (Fig. 9) were Max (100%), or the maximum value of the valence, and three negative emotions: sadness (77%, equaling the median Me of the valence), anger (73%), and fear (72%). In the Spanish press there was never during the course of the volcanic eruption a high confidence or 'trust' as there was in the foreign press (falling from a 92% for foreign articles to 39% for Spanish articles). This was also observed for the joy variable (falling from 84% for foreign articles to 25% for Spanish articles).

4.2.3. Logistic regression classification of press articles

Similarly, a logistic regression analysis was conducted with the Spanish articles by sorting them into group 0 (September and October) and group 1 (November and December). The chi-square goodness-of-fit test resulted in a p-value equal to 0.1870, which allowed us to conclude that with a confidence level of 95% the logistic function adequately fitted to the observed data. Therefore, the relationship between the predictor variables and the group was established by the following equation:

$$G_{Spanish} = \frac{e^{x}}{1 + e^{x}}$$

with x being equal to:

$$x = 0.1423 + 0.1072\ anger + 0.0325\ anticipation + 0.3783\ disgust - 0.3228\ fear + 0.3374\ joy + 0.6342\ sadness - 0.4459\ surprise - 0.0387\ trust - 0.1280\ negative - 0.0373\ positive + 1.5798\ Min + 0.4463\ Q_1 + 5.7468\ Me - 14.8267\ \overline{x} + 8.2243\ Q_3 + 3.2159\ Max$$

4.3. Analysis of the differences between articles written in English and Spanish with a probabilistic neural network

The differences in emotions and sentiments between the articles in the two languages may be appreciated in Table 5. Regarding the eight emotions analyzed, there were significant differences in the emotions expressed in the written stories, with the exception of trust. The median of anger, anticipation, disgust, fear, joy, sadness, and surprise was significantly higher for the articles in English than for those in Spanish. The Cumbre Vieja volcano eruption was experienced with a greater emotional content by foreign journalists than by Spanish journalists. Likewise, it was found that there were no significant differences in the medians of positive emotions between the two groups of articles. However, the medians of the negative sentiments differed significantly, with the medians of English articles higher than the medians of the Spanish articles.

The latter finding indicated that for the foreign journalists sentiments were more negative than for the Spanish journalists. Possible explanations will be discussed later. Fig. 10 illustrates the probabilistic neural network designed to classify the negative/positive news according to the language in which they were written. The percentage of correctly classified training cases was equal to 78.48%, showing in Fig. 11 the classification graph. Note how the region defined by the prescriptor variables was split into two areas where press articles written in Spanish and in English became classified.

4.4. Analysis of the Fourier patterns in the narrative plot

In our study there appeared several characteristic patterns in the temporal change of sentiments and emotions along the narrative plot. Fig. 12 shows examples of the different classes of Fourier plots. In Fig. 12a we may observe how an article starts the story reflecting positive emotions but concludes with words expressing negative emotions. In sentiment analysis parlance, the article has been written with a theatrical style typical of a 'tragedy'. In the present study, we have referred to this pattern as 'Positive'. On the other hand, in Fig. 12b the narrative style is the opposite. That is, a 'Negative' article starts expressing negative emotions but ends with positive emotions, which is typical in theatrical language of the 'comedy'. Likewise, there are articles (Fig. 12c and d) in which positive and negative emotions and sentiments alternate throughout the narrative time. In this case, and regardless of how the article ends, we have studied two kinds of patterns. In particular, we found press articles which begin by expressing negative emotions (Fig. 12c) and articles which start with a predominance of positive emotions (Fig. 12d); they were denominated as 'O_Negative' and 'O_Positive' respectively.

The chi-square statistical analysis of independence shows that there was a significant relationship between the four Fourier plot classes (Positive, Negative, O_Negative, O_Positive) regarding the variation of valence (see Fig. 12), as well as the narrative time and the month in which the article was published (Fig. 13). Thus, both in the articles written in English and those written in Spanish the test of independence allowed us to conclude that there was a dependence (p-value equal to 0.000) between these two factors, i.e. the article

Table 4
Classification table of Spanish press articles.

Actual month	Size	Predicted Group	
		Group 0	**Group 1**
Group 0	40	30 (75.00%)	10 (25.00%)
Group 1	47	12 (25.53%)	35 (74.47%)

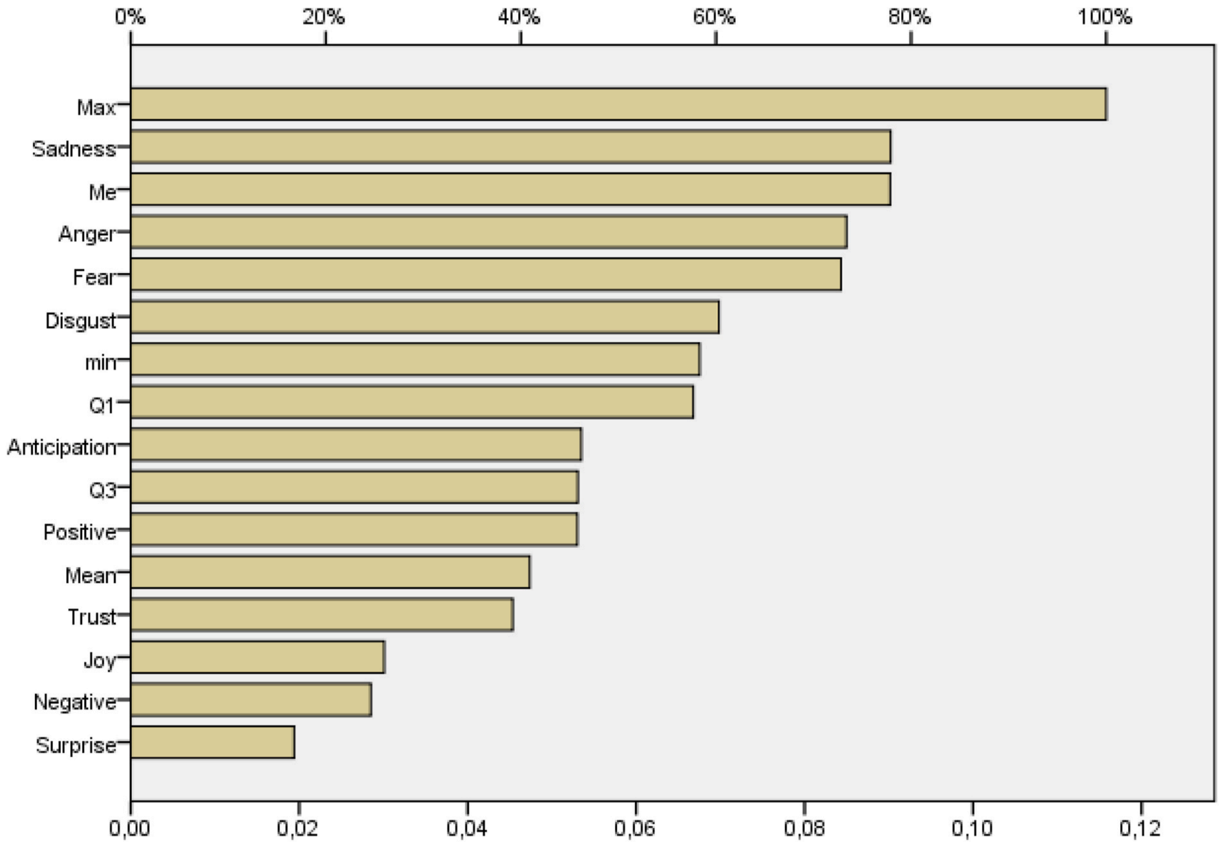

Fig. 9. Bar chart showing the relevance of the predictor variables in the perceptron neural network when classifying the news written in Spanish into two Groups (0 and 1).

Table 5
Contrast of medians between articles written in English and Spanish. Mann-Whitney (Wilcoxon) test.

	Anger	Anticipation	Disgust	Fear	Joy	Sadness	Surprise	Trust	Positive	Negative
p-values	0.0000	0.0106	0.0000	0.0000	0.0020	0.0014	0.0000	0.8224	0.5553	0.0000

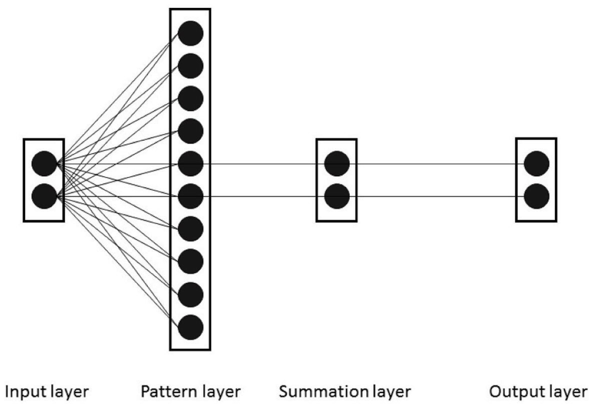

Fig. 10. Probabilistic neural network designed to classify the news in two clusters: 'English' for the foreign newspapers and 'Spanish' for the local press. The input layer has only 2 neurons that receive the value of positive and negative sentiments, being the only two descriptor variables coming from the sentiment analysis. The pattern layer is composed of 10 neurons which calculate the contribution of positive and negative sentiments to the density function of each language group (news written in English or in Spanish). The sum layer assigns a press article to a language group, sending the decision to the output layer.

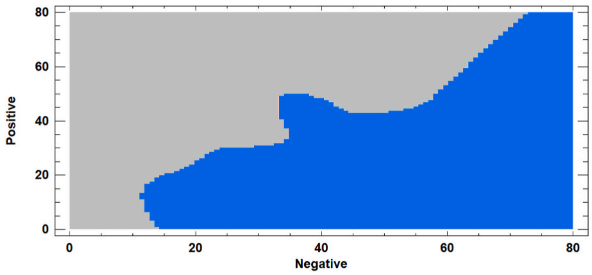

Fig. 11. Classification graph of the probabilistic neural network. The region defined by the prescriptor variables is split into two areas that turn out to be the classified samples: press articles written in Spanish (gray area) and in English (blue area).

month of publication and the Fourier plot pattern.

Fig. 13 shows the journal articles published each month with a given Fourier pattern along the volcanic eruption. In the articles written in English (Fig. 13 Left) the number of 'Positive' articles decreases with the passing months. Likewise, for the 'O_Negative' pattern those articles that begin with a negative paragraph and have an oscillating valence, we found an increase in the number of articles month by month, with a slight decrease at the end of the volcanic eruption. This trend was also observed in the news written in Spanish. Similarly, in the Spanish news (Fig. 13 Right) we observed an increasing number of articles with a negative pattern, the 'Negative' class, month by month. While in the press articles written in English the 'O_Positive' articles repeat the trend observed in the articles with the 'O_Negative' pattern, we have not observed this same trend in the news written in Spanish, more irregularly distributed.

4.5. Regression model between the area occupied by lava and the variables resulting from the sentiment analysis

The multiple linear regression method applied to the English article group allowed us to confirm, with an R^2 equal to 63.04%, the existence of a statistical relationship between the surface (hectares) occupied by the lava (*Lavaflow*) and the variables resulting from sentiment analysis:

$$Lavaflow = 1086.15 - 35.26\ anger + 1.26\ anticipation - 53.05\ \ disgust + 32.28\ \ fear + \\ 54.89\ \ joy + 18.47\ sadness - 35.24\ \ surprise + 15.59\ \ trust - 26.39\ \ negative + \\ 8.06\ \ positive - 78.58\ \ \bar{x}$$

However, when the Spanish local news data were fitted to this model, R^2 was reduced to 43.29%. In spite of this lower value, a similar regression model could also be built.

Lavaflow = 1045.27–29.43 anger +8.68 anticipation +46.37 disgust - 34.67 fear- 1.92 joy +26.66 sadness - 29.19 surprise +7.63 trust - 0.10 negative −7.60 positive +7.97 $\bar{x}$.

Obviously, the model should not be interpreted as a causal relationship, but as a statistical correlation between the area occupied by the lava flow and the predictor variables. Thus, given a date and an article published on such date, the model is able to predict from the predictor variables of sentiment analysis an estimate of the hectares occupied by lava.

5. Discussion

Our analytic work follows a growing trend. Currently the proliferation of programs and libraries on machine learning is such that the application of these techniques has become popular in many disciplines and practical situations, including disaster analysis and risk management [64]. Moreover, the fact that during the 21st century the number of natural disasters related to climate-change has almost doubled [65], and the alarming international events such as the COVID-19 pandemic [66], together have made these new techniques to be present in many studies. In previous papers [47,51,52] we have already successfully applied the combination of sentiment analysis techniques with multivariate analysis methods, e.g., cluster analysis, principal component analysis, and discriminant analysis. Indeed, this methodology is fairly oriented to data analysis when data themselves are texts.

In the results herein obtained, we have observed that the origin of the text, whether the journalist was foreign or domestic, decisively influences the results of sentiment analysis. That the foreign journalists' sentiments were more negative than those of the Spanish journalists demands further discussion. A possible explanation could be that local journalists were more knowing and familiar with the volcanic nature and geology of the Canary Islands and therefore more aware of the actual risk level. Or conversely, it may indicate a higher level of geological knowledge about the potential risks by foreign journalists (as some foreign news were coming from specialized sources), and a higher level of geological interest as well, or perhaps there was a higher proclivity to sensationalize. Since the current data do not allow us to discern between them, or to discard some other possible options, we leave this interesting aspect open for future guesses.

Another interesting result is that the classifiers were less efficient in classifying the news in Spanish than the news written in English. In particular, the discriminant analysis of the English articles reached a 69.01% (59.77% for the Spanish articles) and 76.06% (74.71% in the Spanish articles) depending on whether the periods in which the news were classified were four or two, respectively. Likewise, the perceptron neural network correctly classified in two groups a 78% of the English articles whereas for Spanish-language news the success rate was a 69.0%.

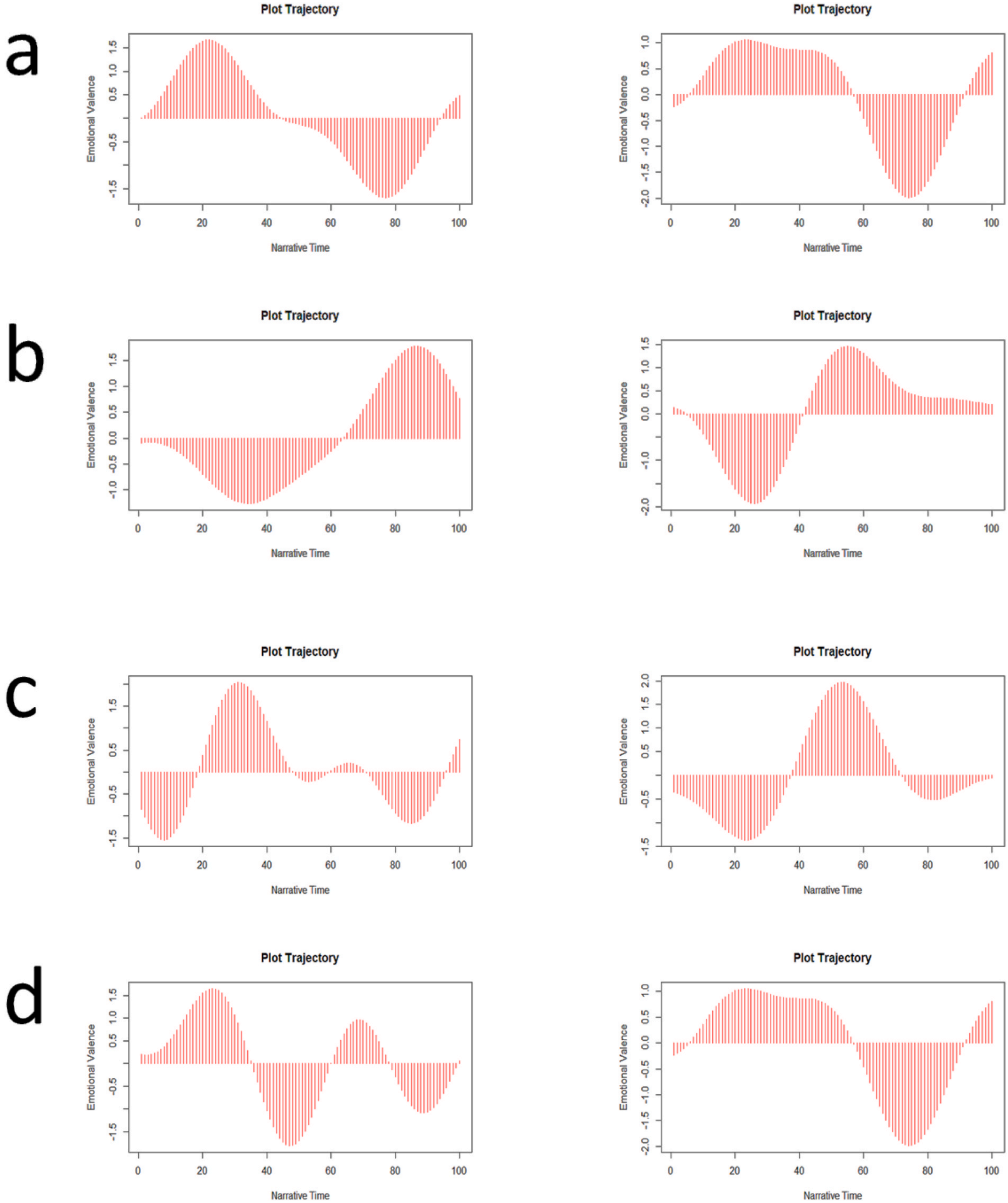

Fig. 12. The elementary Fourier plot patterns. (a) A 'Positive' pattern corresponds to a text that begins positively and ends negatively, as occurs in the theatrical genre of 'tragedy'. (b) A 'Negative' pattern exhibits a change of valences during the narrative time that is the reverse of (a). This pattern is typical of the theatrical genre of 'comedy'. (c) A 'O_Negative' pattern starts negatively and alternates the narrative style between negative-positive valence. (d) A 'O-Positive' pattern starts positively and alternates positive-negative valence.

Further, the medians of sentiments were significantly lower for Spanish journalists than for foreign journalists during the eruption. This would denote a lower predictive power of emotions expressed in the texts of articles written in Spanish by local and national journalists. This fact may explain why the classifiers were less successful with the articles published in Spanish.

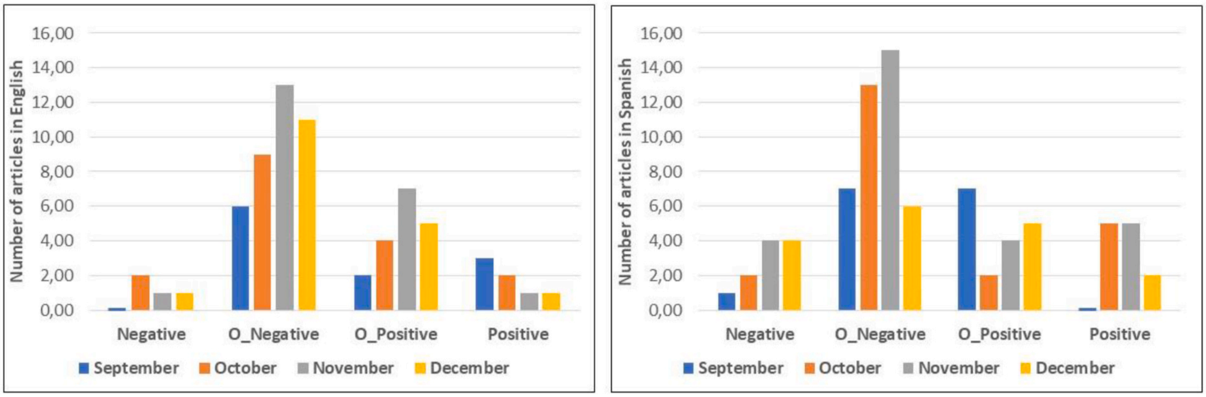

Fig. 13. Number of articles with a given Fourier pattern in the news published in English (Left plot) and Spanish (Right plot).

The fitting of the data to a logistic regression model was also sensitive to the findings discussed above. The fit of the news written in English is better than for news written in Spanish. This fact is not only remarkable because of a higher *p*-value for the news in English, but also if we consider the expression of x in the logistic function. That is, while for news in English x captures only seven emotions (disgust, anger, anticipation, fear, joy, sadness, trust) and the two sentiments (positive, negative), for the news written in Spanish x requires the sixteen variables (eight emotions, two sentiments, six sample statistics of the sentiment vector) resulting from the sentiment analysis.

It was also interesting the presence of a significant relationship between the month of publication of a press article and the type of Fourier plot pattern (variation of valence with respect to the narrative time). Thus, the pattern that the change in valence follows with respect to narrative time depends on the month in which an article was published. We also found that the values of the variables resulting from sentiment analysis could predict the surface area (hectares) occupied by lava. Obviously, this is a statistical or stochastic relationship, not a cause-effect relationship, which is reminiscent of the curious experiment where crickets are 'used' as thermometers [67], predicting the number of chirps the temperature. In any case, there is an interesting psychological analysis on how the different emotions and sentiments correlated with the advancement of lava flow.

In sum, we believe that the results obtained show how press articles can be useful as a source of information to evaluate the social and environmental consequences of a natural disaster or catastrophe. Moreover, the combination of sentiment analysis with multivariate analysis and machine learning techniques – a main originality of our approach – could improve the protocols oriented to the evaluation of the environmental, economic, and social impacts of a natural disaster or catastrophe. Likewise, the described methodology could also allow the evaluation from ancient texts, particularly if there were different sources of historical interest, searching for the impact and consequences of those catastrophes and disasters occurred long ago.

The manifold aftermaths of volcanic eruptions, assessed as natural phenomena occurring in complex socio-environmental frameworks, require integrated approaches to their spatial and temporal dynamics. It makes necessary the application of robust tools adequately coupled with the related social and individual emotional dynamics, in terms of components, connections, scale, and context, as well as the associated space-time frontiers [68]. Governance of catastrophic natural disasters may not leave behind the interpretation of people's sentiments. With that aim, additional research about natural language processing techniques would be necessary, by improving deep learning models, such as convolutional neural networks for local data extraction, recurrent models for dependence ordering in sequential data, or deep ensemble learning, among others [69].

6. Conclusions

The Cumbre Vieja ("Tajogaite") eruption is shown in this work as a valuable example about how the application of novel approaches and the combination of tools may provide new frameworks for understanding people's sentiments and needs under the frame of various types of disaster risks.

The results herein obtained allow us to conclude that sentiment analysis applied to press media is a useful technique that in combination with multivariate statistical methods and machine learning classifiers is able to classify an article into a series of groups or clusters and to establish useful correlations. Thus, the emotions and sentiments expressed in the text of a press article can be closely related to the stage or period in which the article was written during the course of a disaster. Indeed, the emotions and sentiments expressed in a text change over time in response to the direction of events and to the social impact generated. This approach would allow improved designs of tools to study the evolution of social and environmental impacts of a natural phenomenon and to assess how impacts change over time.

Another potentially important fact is the origin (local press *vs* foreign press)of the information source. There might be significant differences in the emotions and sentiments expressed in the text of a given article depending on whether the news analyzed were coming from a foreign or a local media. We have also found that there is a statistical relationship between the emotions and sentiments expressed in an article and the impact of a natural disaster. And we were able to gauge the valence inherent of each individual article across narrative time and in an aggregate cluster.

Essentially, we demonstrate the usefulness of the press media as a source of information complementing currently popular sources such as Twitter and other platforms. As we have argued, when a population is affected by a crisis or a natural disaster, it is important analyzing the three main layers of communication: immediacy and spontaneity (social media), curated press news in newspapers and online (press media), and in-depth reflections and reports (articles and reviews). These three modalities of communication have their own social function during a crisis or disaster, with their respective characteristics, advantages, and inconveniences.

Therefore, the main findings of this research may constitute a helpful resource for developing ameliorated insights into the way a society reacts to volcanic activity and may strengthen the foundations for decision-making under different temporal horizons, also considering that other eruptions may occur in the future.

So, this study's outcomes may contribute to fill in knowledge gaps related to potential information dissemination biases during crises and natural disasters, and to improve the analysis of the disaster risk-reduction strategies followed.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

The authors acknowledge the excellent revision work performed by the anonymous reviewers. Their comments, suggestions, and criticisms have much contributed to improve the quality of this article. The authors are also very grateful to Pedro C. Marijuán for his comments and suggestions that were addressed into the revised version of the article.

References

[1] M.A. Longpré, Reactivation of cumbre vieja volcano, Science (1979), https://doi.org/10.1126/science.abm9423, 374 (2021.
[2] Comisión Mixta, Informe sobre las actuaciones y medidas emprendidas tras la erupción del volcán de Cumbre Vieja (La Palma), seis meses después del inicio de la emergencia, (n.d.) 17–20. https://www.mpr.gob.es/prencom/notas/Documents/2022/060622-informe_palma.pdf (accessed March 6, 2023).
[3] European Union, Copernicus. https://www.copernicus.eu/en.
[4] P.A. Torres-González, N. Luengo-Oroz, H. Lamolda, W. D'Alessandro, H. Albert, I. Iribarren, D. Moure-García, V. Soler, Unrest signals after 46 years of quiescence at cumbre vieja, La Palma, Canary Islands, J. Volcanol. Geoth. Res. 392 (2020), https://doi.org/10.1016/j.jvolgeores.2019.106757.
[5] E. Padrón, N.M. Pérez, P.A. Hernández, H. Sumino, G. Melián, J. Barrancos, D. Nolasco, G. Padilla, Helium emission at Cumbre Vieja volcano, La Palma, Canary Islands, Chem. Geol. (2012) 312–313, https://doi.org/10.1016/j.chemgeo.2012.04.018.
[6] E. Padrón, N.M. Pérez, F. Rodríguez, G. Melián, P.A. Hernández, H. Sumino, G. Padilla, J. Barrancos, S. Dionis, K. Notsu, D. Calvo, Dynamics of diffuse carbon dioxide emissions from Cumbre Vieja volcano, La Palma, Canary Islands, Bull. Volcanol. 77 (2015), https://doi.org/10.1007/s00445-015-0914-2.
[7] S.M. Abadie, J.C. Harris, S.T. Grilli, R. Fabre, Numerical modeling of tsunami waves generated by the flank collapse of the Cumbre Vieja volcano (La Palma, Canary Islands): tsunami source and near field effects, J Geophys Res Oceans 117 (2012), https://doi.org/10.1029/2011JC007646.
[8] J.C. Harris, S.T. Grilli, S. Abadie, T.T. Bakhsh, Near- and far-field tsunami hazard from the potential flank collapse of the Cumbre Vieja volcano, in: Proceedings of the International Offshore and Polar Engineering Conference, 2012.
[9] B. Tehranirad, J.C. Harris, A.R. Grilli, S.T. Grilli, S. Abadie, J.T. Kirby, F. Shi, Far-field tsunami impact in the north atlantic basin from large scale flank collapses of the Cumbre Vieja volcano, La Palma, Pure Appl. Geophys. 172 (2015), https://doi.org/10.1007/s00024-015-1135-5.
[10] A.M.G. Hernández, Refugiados en tierra propia. A Propósito de la situación de "duelo social" ocasionado por el volcán de la Cumbre Vieja en La Palma, Revista Ene de Enfermería. 15 (2021).
[11] A. Geyer, M. Aulinas, H. Albert, A.M. Álvarez-Valero, L'impacte del vulcanisme en la història de la humanitat, ATZAVARA 33 (2023) 11–26.
[12] D. Kahneman, Thinking, Fast and Slow, Farrar, Straus and Giroux, New York, 2011.
[13] K.C. Yang, O. Varol, C.A. Davis, E. Ferrara, A. Flammini, F. Menczer, Arming the public with artificial intelligence to counter social bots, Hum Behav Emerg Technol 1 (2019), https://doi.org/10.1002/hbe2.115.
[14] C. Zhang, Y. Yang, A. Mostafavi, Revealing Unfairness in social media contributors' attention to vulnerable urban areas during disasters, Int. J. Disaster Risk Reduc. 58 (2021), https://doi.org/10.1016/j.ijdrr.2021.102160.
[15] V. Balakrishnan, W.Z. Ng, M.C. Soo, G.J. Han, C.J. Lee, Infodemic and fake news – a comprehensive overview of its global magnitude during the COVID-19 pandemic in 2021: a scoping review, Int. J. Disaster Risk Reduc. 78 (2022), 103144, https://doi.org/10.1016/j.ijdrr.2022.103144.
[16] H. Allcott, M. Gentzkow, Social media and fake news in the 2016 election, J. Econ. Perspect. 31 (2017), https://doi.org/10.1257/jep.31.2.211.
[17] S. Bashir, S. Bano, S. Shueb, S. Gul, A.A. Mir, R. Ashraf, Shakeela, N. Noor, Twitter chirps for Syrian people: sentiment analysis of tweets related to Syria Chemical Attack, Int. J. Disaster Risk Reduc. 62 (2021), https://doi.org/10.1016/j.ijdrr.2021.102397.
[18] M.O. Badmus, When the storm is over: sentiments, communities and information flow in the aftermath of Hurricane Dorian, Int. J. Disaster Risk Reduc. 47 (2020), 101645, https://doi.org/10.1016/j.ijdrr.2020.101645.
[19] M. Cassidy, L. Mani, Huge volcanic eruptions: time to prepare, Nature 608 (2022) 469–471, https://doi.org/10.1038/d41586-022-02177-x.
[20] H. Sigurdsson, R.M.C. Lopes, Volcanoes in literature and film, in: The Encyclopedia of Volcanoes, 2015, https://doi.org/10.1016/b978-0-12-385938-9.00078-x.
[21] M. Nordvig, Volcanoes in Old Norse Mythology, Arc Humanities Press, 2021, https://doi.org/10.2307/j.ctv1k76ht8.
[22] J. Barclay, K. Haynes, B. Houghton, D. Johnston, Social processes and volcanic risk reduction, in: The Encyclopedia of Volcanoes, 2015, https://doi.org/10.1016/b978-0-12-385938-9.00069-9.
[23] L.F. Barrett, Solving the emotion paradox: categorization and the experience of emotion, Pers. Soc. Psychol. Rev. 10 (2006) 20–46, https://doi.org/10.1207/s15327957pspr1001_2.
[24] J.A. Russell, L.F. Barrett, Core affect, prototypical emotional episodes, and other things called emotion: dissecting the elephant, J. Pers. Soc. Psychol. 76 (1999) 805–819, https://doi.org/10.1037/0022-3514.76.5.805.
[25] P. Ekman, Basic emotions, in: T. Dalgleish, M. Power (Eds.), Handbook of Cognition and Emotion, John Wiley & Sons, Sussex (UK), 1999, pp. 45–60.
[26] J. Navarro, P.C. Marijuán, The natural, artificial, and social domains of intelligence: a triune approach, in: IS4SI 2021, MDPI, Basel Switzerland, 2022, p. 2, https://doi.org/10.3390/proceedings2022081002.
[27] V. Csany, Contribution of the genetical and neural memory to animal intelligence, in: H. Jerison, I. Jerison (Eds.), Intelligence and Evolutionary Biology, Ecological Sciences, Springer-Verlag, Berlin, 1988.

[28] M.R. Dove, Perception of volcanic eruption as agent of change on Merapi volcano, Central Java, J. Volcanol. Geoth. Res. 172 (2008), https://doi.org/10.1016/j.jvolgeores.2007.12.037.
[29] S. Bachri, J. Stötter, M. Monreal, J. Sartohadi, The calamity of eruptions, or an eruption of benefits? Mt. Bromo human-volcano system a case study of an open-risk perception, Nat. Hazards Earth Syst. Sci. 15 (2015), https://doi.org/10.5194/nhess-15-277-2015.
[30] J.C. Gomez-Zapata, C. Parrado, T. Frimberger, F. Barragán-Ochoa, F. Brill, K. Büche, M. Krautblatter, M. Langbein, M. Pittore, H. Rosero-Velásquez, E. Schoepfer, H. Spahn, C. Zapata-Tapia, Community perception and communication of volcanic risk from the cotopaxi volcano in Latacunga, Ecuador, Sustainability (2021) 13, https://doi.org/10.3390/su13041714.
[31] C. Clavel, Z. Callejas, Sentiment analysis: from opinion mining to human-agent interaction, IEEE Trans Affect Comput 7 (2016), https://doi.org/10.1109/TAFFC.2015.2444846.
[32] C.C. Chan, K.J. Liszka, Application of Rough Set Theory to Sentiment Analysis of Microblog Data, vol. 43, Intelligent Systems Reference Library, 2013, https://doi.org/10.1007/978-3-642-30341-8_10.
[33] G. Vinodhini, R. Chandrasekaran, Sentiment analysis and opinion mining : a survey international journal of advanced research in sentiment analysis and opinion mining : a, Int. J. Adv. Res. Comput. Sci. Software Eng. 2 (2012).
[34] E. Şentürk, M.A. Adil, M. Toker, M.C. Iban, Ö. Akyürek, Long-term ion hydration process and lithosphere-atmosphere coupling following the 2021 Fagradalsfjall volcanic eruption using remotely sensed data, Adv. Space Res. (2022), https://doi.org/10.1016/j.asr.2022.12.008.
[35] S. D'Arcangelo, A. Bonforte, A. de Santis, S.R. Maugeri, L. Perrone, M. Soldani, G. Arena, F. Brogi, M. Calcara, S.A. Campuzano, G. Cianchini, A. del Corpo, D. di Mauro, C. Fidani, A. Ippolito, S. Lepidi, D. Marchetti, A. Nardi, M. Orlando, A. Piscini, M. Regi, D. Sabbagh, Z. Zhima, R. Yan, A multi-parametric and multi-layer study to investigate the largest 2022 Hunga Tonga–Hunga Ha'apai eruptions, Rem. Sens. 14 (2022) 3649, https://doi.org/10.3390/rs14153649.
[36] P. Lynett, M. McCann, Z. Zhou, W. Renteria, J. Borrero, D. Greer, O. Fa'anunu, C. Bosserelle, B. Jaffe, S. la Selle, A. Ritchie, A. Snyder, B. Nasr, J. Bott, N. Graehl, C. Synolakis, B. Ebrahimi, G.E. Cinar, Diverse tsunamigenesis triggered by the Hunga Tonga-Hunga Ha'apai eruption, Nature 609 (2022) 728–733, https://doi.org/10.1038/s41586-022-05170-6.
[37] L. Ye, H. Kanamori, L. Rivera, T. Lay, Y. Zhou, D. Sianipar, K. Satake, The 22 December 2018 tsunami from flank collapse of Anak Krakatau volcano during eruption, Sci. Adv. 6 (2020), https://doi.org/10.1126/sciadv.aaz1377.
[38] J.C. Carracedo, V.R. Troll, J.M.D. Day, H. Geiger, M. Aulinas, V. Soler, F.M. Deegan, F.J. Perez-Torrado, G. Gisbert, E. Gazel, A. Rodriguez-Gonzalez, H. Albert, The 2021 eruption of the Cumbre Vieja volcanic ridge on La Palma, Canary Islands, Geol. Today 38 (2022) 94–107, https://doi.org/10.1111/gto.12388.
[39] P.J. González, Volcano-tectonic control of Cumbre Vieja, Science 375 (1979) 1348–1349, https://doi.org/10.1126/science.abn5148, 2022.
[40] F. Yuan, M. Li, R. Liu, Understanding the evolutions of public responses using social media: hurricane Matthew case study, Int. J. Disaster Risk Reduc. 51 (2020), 101798, https://doi.org/10.1016/j.ijdrr.2020.101798.
[41] V.K. Neppalli, C. Caragea, A. Squicciarini, A. Tapia, S. Stehle, Sentiment analysis during Hurricane Sandy in emergency response, Int. J. Disaster Risk Reduc. 21 (2017) 213–222, https://doi.org/10.1016/j.ijdrr.2016.12.011.
[42] S. Behl, A. Rao, S. Aggarwal, S. Chadha, H.S. Pannu, Twitter for disaster relief through sentiment analysis for COVID-19 and natural hazard crises, Int. J. Disaster Risk Reduc. 55 (2021), 102101, https://doi.org/10.1016/j.ijdrr.2021.102101.
[43] G. Bhullar, A. Khullar, A. Kumar, A. Sharma, H.S. Pannu, A. Malhi, Time series sentiment analysis (SA) of relief operations using social media (SM) platform for efficient resource management, Int. J. Disaster Risk Reduc. 75 (2022), 102979, https://doi.org/10.1016/j.ijdrr.2022.102979.
[44] M. Berardi, L. Santamaria Amato, F. Cigna, D. Tapete, M. Siciliani de Cumis, Text mining from free unstructured text: an experiment of time series retrieval for volcano monitoring, Appl. Sci. 12 (2022) 3503, https://doi.org/10.3390/app12073503.
[45] J.R. Ragini, P.M.R. Anand, V. Bhaskar, Big data analytics for disaster response and recovery through sentiment analysis, Int. J. Inf. Manag. 42 (2018) 13–24, https://doi.org/10.1016/j.ijinfomgt.2018.05.004.
[46] S. Jabalameli, Y. Xu, S. Shetty, Spatial and sentiment analysis of public opinion toward COVID-19 pandemic using twitter data: at the early stage of vaccination, Int. J. Disaster Risk Reduc. 80 (2022), 103204, https://doi.org/10.1016/j.ijdrr.2022.103204.
[47] A. Turón, A. Altuzarra, J.M. Moreno-Jiménez, J. Navarro, Evolution of social mood in Spain throughout the COVID-19 vaccination process: a machine learning approach to tweets analysis, Publ. Health 215 (2023) 83–90, https://doi.org/10.1016/j.puhe.2022.12.003.
[48] D. Contreras, S. Wilkinson, N. Balan, P. James, Assessing post-disaster recovery using sentiment analysis: The case of L'Aquila, Italy, Earthquake Spectra 38 (2022) 81–108, https://doi.org/10.1177/87552930211036486.
[49] S. Behl, A. Rao, S. Aggarwal, S. Chadha, H.S. Pannu, Twitter for disaster relief through sentiment analysis for COVID-19 and natural hazard crises, Int. J. Disaster Risk Reduc. 55 (2021), 102101, https://doi.org/10.1016/j.ijdrr.2021.102101.
[50] J. Rexiline Ragini, P.M. Rubesh Anand, V. Bhaskar, Mining crisis information: a strategic approach for detection of people at risk through social media analysis, Int. J. Disaster Risk Reduc. 27 (2018) 556–566, https://doi.org/10.1016/j.ijdrr.2017.12.002.
[51] R. Lahoz-Beltra, C.C. López, LENNA (Learning Emotions Neural Network Assisted): an empathic chatbot designed to study the simulation of emotions in a bot and their analysis in a conversation, Computers 10 (2021) 170, https://doi.org/10.3390/computers10120170.
[52] J. Navarro, A. Turón, A. Altuzarra, R. Lahoz-Beltra, Comparative sentiment analysis of COVID-19: a machine learning approach, in: 7th ICDSST 2021 (International Conference on Decision Support System Technology), Loughborough (Reino Unido), 2021.
[53] M.L. Jockers, Syuzhet Release 1.0.6. , (n.d.). https://Cran.r-Project.Org/Web/Packages/Syuzhet/Syuzhet.Pdf (accessed March 6, 2023).
[54] F. Bravo-Marquez, E. Frank, S.M. Mohammad, B. Pfahringer, IEEE/WIC/ACM International Conference on Web Intelligence (WI), in: Determining Word-Emotion Associations from Tweets by Multi-Label Classification, IEEE, 2016, pp. 536–539, https://doi.org/10.1109/WI.2016.0091, 2016.
[55] S.M. Mohammad, P.D. Turney, Crowdsourcing a word-emotion association lexicon, Comput. Intell. 29 (2013) 436–465, https://doi.org/10.1111/j.1467-8640.2012.00460.x.
[56] S. Mohammad. https://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm. (Accessed 6 March 2023). https://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm.
[57] M.L. Jockers, Revealing Sentiment and Plot Arcs with the Syuzhet Package, 2015. https://www.matthewjockers.net/2015/02/02/syuzhet/. (Accessed 6 March 2023).
[58] A.J. Izenman, Modern Multivariate Statistical Techniques, Springer New York, New York, NY, 2008, https://doi.org/10.1007/978-0-387-78189-1.
[59] R.A. Fisher, The use of multiple measurements in taxonomic problems, Ann Eugen 7 (1936) 179–188, https://doi.org/10.1111/j.1469-1809.1936.tb02137.x.
[60] R. Lahoz-Beltra, Bioinformática: Simulación, Vida Artificial e Inteligencia Artificial, Ediciones Díaz de Santos, Madrid, Spain, 2004.
[61] M.I. Hidayat Jati, Suroso, P.B. Santoso, Prediction of flood areas using the logistic regression method (case study of the provinces Banten, DKI Jakarta, and West Java), J Phys Conf Ser (2019), https://doi.org/10.1088/1742-6596/1367/1/012087.
[62] R. Muñoz-Mas, S. Fukuda, J. Pórtoles, F. Martínez-Capel, Revisiting probabilistic neural networks: a comparative study with support vector machines and the microhabitat suitability for the Eastern Iberian chub (Squalius valentinus), Ecol. Inf. 43 (2018), https://doi.org/10.1016/j.ecoinf.2017.10.008.
[63] B. Mohebali, A. Tahmassebi, A. Meyer-Baese, A.H. Gandomi, Probabilistic neural networks: a brief overview of theory, implementation, and application, in: P.D. T.B.S.C.R.C. Samui (Ed.), Handbook of Probabilistic Models, Butterworth-Heinemann, Oxford, UK, 2020, pp. 347–367.
[64] V. Linardos, M. Drakaki, P. Tzionas, Y.L. Karnavas, Machine learning in disaster management: recent developments in methods and applications, Mach Learn Knowl Extr 4 (2022) 446–473, https://doi.org/10.3390/make4020020.
[65] G. Alimonti, L. Mariani, On the Exponential Increase of Natural Disasters in the 20th Century. Science, climat et énergie, (n.d.). https://www.science-climat-energie.be/on-the-exponential-increase-of-natural-disasters-in-the-20th-century/(accessed March 6, 2023).
[66] World Health Organization, Coronavirus Disease (COVID-19) Pandemic, (n.d.). https://www.Who.Int/Emergencies/Diseases/Novel-Coronavirus-2019 (accessed March 6, 2023).

[67] A.E. Dolbear, The cricket as a thermometer, Am. Nat. 31 (1897), https://doi.org/10.1086/276739.
[68] S. Pulver, N. Ulibarri, K.L. Sobocinski, S.M. Alexander, M.L. Johnson, P.F. McCord, J. Dell'Angelo, Frontiers in socio-environmental research: components, connections, scale, and context, Ecol. Soc. 23 (2018), https://doi.org/10.5751/ES-10280-230323 art23.
[69] A. Mohammadi, A. Shaverizade, Ensemble deep learning for aspect-based sentiment analysis, Int. J. Nonlinear Anal. Appl. 12 (2021), https://doi.org/10.22075/IJNAA.2021.4769.

7. DISCUSIÓN

7. DISCUSIÓN

La línea de investigación que sirve de base a esta Tesis, realizada mediante un compendio de publicaciones, está centrada en la dinámica social de las emociones y en los patrones emocionales colectivos analizados mediante técnicas de inteligencia artificial. Tanto el concepto de sociotipo como el de inteligencia natural nos han ayudado a clarificar el encaje de las emociones con las relaciones sociales, así como con el mantenimiento de los lazos sociales a lo largo del tiempo. Lo cual nos permite conectar estas ideas con propuestas interesantes para la investigación en *sentiment analysis*.

Primero, en el Artículo A, destacamos algunas importantes consecuencias (sociales, económicas y culturales) derivadas de dichas dinámicas emocionales. Estas incluyen la ausencia de relaciones y la desaparición de la red de lazos sociales o sociotipo del individuo. Los problemas derivados del creciente aislamiento y soledad de los mayores constituyen una amenaza de primera magnitud para el sistema público de salud y la asistencia social. Su detección mediante herramientas simplificadas al alcance inmediato de especialistas geriátricos o de generalistas de atención primaria resulta fundamental.

Hemos desarrollado una nueva versión del cuestionario objetivo de cuatro ítems con dos opciones de respuesta, extrayendo cuatro de los 12 ítems originales y dicotomizando las seis opciones de respuesta originales del SOCQ. Este cuestionario SOCG-4 abreviado para población geriátrica se probado en un estudio piloto (véase Apéndice A de (Navarro et al., 2022)).

En comparación con otras pruebas similares, éstas implican un sentido de dependencia, vulnerabilidad y necesidad de apoyo externo para satisfacer las necesidades personales, incluso suponiendo un estado depresivo. En cambio, el sociotipo se refiere a conversaciones espontáneas y a un sentido de empoderamiento mientras el sujeto gestiona sus relaciones de forma autónoma. Los sujetos responden con mayor facilidad y buen humor a las preguntas del SOCQ y del SOCG-4, en particular a la pregunta sobre la risa. La facilidad de uso ha sido validada por trabajadores sociales que ya utilizan institucionalmente el cuestionario SOCQ, en comparación con UCLA, y por el uso en el trabajo de campo piloto del SOCG-4 descrito en el Apéndice A. Los resultados obtenidos en el trabajo de campo y los comentarios de los entrevistadores confirmaron la facilidad, utilidad y capacidad discriminatoria del nuevo cuestionario.

Segundo, en el Artículo B, sobre la aproximación teórica a las emociones sociales, queremos enfatizar el interés investigador del concepto antes comentado del *sociotipo*. En efecto, los constructos genotipo y fenotipo para la especie humana, a pesar de sus respectivos grados de variabilidad, irían acompañados de una métrica del sociotipo, representativa de la relativa constancia del entorno social básico al que estarían evolutivamente adaptados los individuos de nuestra especie.

Los dominios relacionales diferenciados dentro del sociotipo son importantes para un enfoque más matizado de las emociones sociales. Destacamos nuestra distinción entre emociones primarias para el círculo relacional más cercano ("vínculos fuertes"), emociones secundarias para las relaciones interindividuales ("vínculos débiles") y, por último, unas hipotéticas emociones terciarias para el ámbito ultrasocial. En conjunto, existe una elegancia de diseño y una economía evidente en la correspondencia evolutiva entre los grupos de emociones y las clases de vínculos sociales.

Entonces, como estrategia de investigación, en lugar de buscar clasificaciones de emociones más sofisticadas o listas cada vez más largas, nos hemos centrado en los patrones contextuales que guían o desencadenan nuestro ajuste socioemocional. Los ejes de patrones interindividuales y ultrasociales propuestos son hipótesis provisionales, pero a través del *bootstrap* del análisis de sentimientos podríamos empezar a materializar la interrelación dialéctica de los patrones o contextos y las distintas emociones. A partir de ahí, podríamos considerar el refinamiento de las clasificaciones y la ampliación de las listas de emociones bajo una nueva perspectiva.

Nuestro enfoque final del análisis de sentimientos dentro de la inteligencia artificial actual ha demostrado que podría existir una conexión eficaz entre los programas de investigación de las dos ramas de la inteligencia: la natural y la artificial, la IN y la IA. Creemos que hoy día hay suficiente potencial de investigación para aplicar el músculo de la IA a las emociones sociales y sus circunstancias desencadenantes. En la dicotomía entre la inteligencia natural frente a inteligencia artificial, bien podemos responder: lo que estaba separado debe volver a unirse.

En tercer lugar, Artículo C, en lo que respecta al análisis de sentimientos para evaluar el estado anímico durante el proceso de vacunación contra el COVID-19, el estudio ha obtenido una serie de resultados congruentes en cuanto a las redes sociales implicadas, la evolución del estado anímico social unida a la dinámica de estas redes y el análisis del sentimiento representado en los gráficos de trayectoria. Esta congruencia general entre los distintos tipos de resultados obtenidos puede interpretarse como un aspecto muy prometedor del enfoque que hemos desarrollado.

Señalemos primeramente que, en lo que respecta a la evolución de las redes sociales representadas, la dinámica de agrupación durante las cuatro fases de vacunación diferenciadas es sorprendentemente precisa, captando fielmente la evolución de la opinión pública durante el proceso de vacunación. El análisis de la red de retweets no sólo muestra las interconexiones y la agrupación de la comunidad de tuiteros en torno a grupos de interés, sino que también muestra cómo la estructura de estos grupos varía a lo largo del proceso. Se observa cómo las decisiones de salud pública y otras circunstancias del entorno que provocan los cambios de ánimo no sólo se reflejan en la forma en que se agrupan los tuiteros, sino también en quiénes son sus referentes a la hora de compartir información.

Se ha mostrado la importancia de la comunicación de salud pública desde fuentes oficiales, ya que sus retweets de otros usuarios pueden llegar a muchas más personas que no siguen las cuentas oficiales. Esto representa una estrategia de comunicación rentable para promover la salud pública. En este sentido, podemos observar que la mayoría de los líderes políticos internacionales están recurriendo cada vez más a las redes sociales para difundir información sobre pandemias, planes de respuesta, medidas de salud pública y la conexión con los ciudadanos. Esto implica una serie de elecciones estratégicas para adoptar un enfoque más positivo con el fin de influir en la opinión y la acción, y fomentar el cumplimiento de las normas y estándares de salud pública.

Por último, el análisis de sentimientos ofrece una panoplia creciente de nuevas herramientas y paradigmas para explicar la aparición de estados de ánimo sociales y fenómenos de contagio emocional, que son tan importantes en nuestras sociedades, incluida la actual "epidemia de soledad” (Bernal, 2013; Navarro et al., 2022).

En cuarto lugar, sobre el Artículo D, cabe enmarcar nuestro trabajo analítico en una tendencia creciente de la aplicación del análisis de sentimientos para las catástrofes naturales y la gestión de riesgos. Además, el hecho de que durante el siglo XXI casi se haya duplicado el número de desastres naturales relacionados con el cambio climático y eventos internacionales como la propia pandemia de COVID ha llevado a que estas nuevas técnicas estén presentes en numerosos estudios. Aquí hemos aplicado con éxito la combinación de técnicas de análisis de sentimientos con métodos de análisis multivariante (análisis de conglomerados, análisis de componentes principales y análisis discriminante). Y lo hemos aplicado al análisis de noticias de prensa en un sentido amplio. De hecho, la metodología seguida está particularmente orientada al análisis de datos cuando los datos en sí son textos.

En los resultados aquí obtenidos, hemos observado que el origen del texto, ya sea de un periodista extranjero o nacional, influye de manera decisiva en los sentimientos expresados. Curiosamente, el hecho de que los sentimientos de los periodistas extranjeros fueran más negativos que los de los periodistas españoles nos ha llevado a una interesante discusión adicional. Otro resultado curioso es que los clasificadores fueron menos eficientes a la hora de clasificar las noticias en español que las escritas en inglés.

También resultó interesante la presencia de una relación significativa entre el mes de publicación de un artículo de prensa y el tipo de patrón del gráfico de Fourier (variación de la valencia con respecto al tiempo narrativo). También comprobamos que los valores de las variables resultantes del análisis de sentimientos podían predecir la superficie (hectáreas) ocupada por la lava. En cualquier caso, existe un análisis psicológico interesante sobre cómo las diferentes emociones y sentimientos se correlacionan con el avance del flujo de lava.

En resumen, creemos que los resultados obtenidos muestran cómo los artículos de prensa pueden ser útiles como fuente de información para evaluar las consecuencias sociales y medioambientales de un desastre natural o catástrofe. Además, la combinación del análisis de sentimiento con técnicas de análisis multivariante y aprendizaje automático, una de las principales originalidades de nuestro enfoque, podría mejorar los protocolos orientados a la evaluación de los impactos ambientales, económicos y sociales de un desastre natural o catástrofe. Asimismo, la metodología descrita también podría permitir la evaluación a partir de textos antiguos, especialmente si existen diferentes fuentes de interés histórico, buscando el impacto y las consecuencias de aquellas catástrofes y desastres ocurridos en el pasado.

Las múltiples secuelas de las erupciones volcánicas, evaluadas como fenómenos naturales que ocurren en marcos socioambientales complejos, requieren enfoques integrados de su dinámica espacial y temporal. La gobernanza de los desastres naturales catastróficos no puede pasar por alto la interpretación de los sentimientos de las personas. Con ese objetivo, sería necesaria una investigación adicional sobre las técnicas de procesamiento del lenguaje natural, mejorando los modelos de aprendizaje profundo, como las redes neuronales convolucionales para la extracción de datos locales, los modelos recurrentes para el ordenamiento de la dependencia en datos secuenciales o el aprendizaje profundo de conjuntos.

8. CONCLUSIONES

8. CONCLUSIONES

Para encabezar estas conclusiones, podemos enfatizar el argumento central de esta Tesis sobre el papel de las emociones en el ámbito social, que es el orientar adaptativamente y dotar de “sentido común” a las diversas situaciones del individuo en el seno del colectivo social, dirigiendo eficazmente sus decisiones y acciones desde un punto de vista adaptativo, complementando así el papel tradicional que se ha atribuido exclusivamente a la “racionalidad”.

En referencia a la ausencia de relaciones sociales, nuestro trabajo pone de manifiesto que la protección más importante para los mayores, es decir, contar con un mínimo de relaciones sociales, suele quedar fuera del alcance de la atención médica actual, que está muy segmentada y centrada en los fallos orgánicos y las enfermedades relacionadas con la edad. Al prestar especial atención al aislamiento social y a la soledad percibida, el médico de atención primaria o el facultativo de geriatría estarán, de hecho, vigilando el estado del factor de riesgo más importante para la salud del paciente.

Enfatizamos que, a partir de ahí, la prescripción de "píldoras de socialización", a través de la coordinación con los centros de día municipales o regionales para los mayores aislados, puede mejorar la calidad de vida y, en algunos casos, evitar la medicación antidepresiva y el síndrome perjudicial de la polifarmacia en los mayores.

En este contexto, lo que significa la iniciativa del sociotipo, y en particular el SOCG-4, es la promoción de un instrumento fácil de tener a mano en la mesa de la atención primaria y la clínica geriátrica, que actúe como recordatorio del principal factor de riesgo para la salud de la persona mayor que se encuentra delante del médico. Podemos insistir en la importancia de esta nueva métrica abreviada dado el aumento de la soledad tras la pandemia de COVID-19.

Pero independientemente de los resultados concretos del SOCG-4 y de sus limitaciones básicas, debería estimular el debate para contribuir a un cambio de mentalidad en el sistema asistencial actual, volviendo a centrarse en las necesidades de socialización, no bien cubiertas en medio de los frecuentes cambios administrativos y tecnológicos. El rápido envejecimiento de las sociedades en todo el mundo exige nuevas orientaciones, nuevas estrategias de intervención y nuevos instrumentos de cribado para las gerociencias. En particular, la sociedad pospandémica necesita urgentemente este tipo de planteamientos.

En la aproximación que proponemos a las emociones sociales, destaca de nuevo el interés investigador del sociotipo. Los dominios relacionales diferenciados dentro del sociotipo son importantes para un mejor entendimiento de las emociones sociales. Nuestra distinción entre emociones primarias para el círculo relacional más cercano ("vínculos fuertes") y emociones secundarias para las relaciones interindividuales ("vínculos débiles") adquiere una indudable relevancia.

Asimismo, los ejes de patrones interindividuales y sociales aquí propuestos podrían ayudar a materializar los contextos versus emociones en su interrelación dialéctica.

Igualmente, nuestro enfoque final del análisis de sentimientos dentro de la inteligencia artificial actual ha demostrado que podría existir una conexión eficaz y eficiente entre los programas de investigación de las dos ramas de la inteligencia: hay suficiente potencial de investigación en la IA para un enfoque más adecuado de las emociones sociales y sus circunstancias desencadenantes.

En cuanto a la aplicación del análisis de sentimientos durante la pandemia, el nuevo enfoque desarrollado ha demostrado su eficacia. Ha combinado técnicas de aprendizaje automático (análisis de sentimientos y minería de datos) con métodos de análisis multivariante (análisis de redes sociales y minería de textos). Y se ha utilizado software libre, de muy fácil acceso y uso. Se pretende integrar todas estas herramientas informáticas en un Sistema de Ayuda a la Decisión, más fácil de utilizar e interpretar los resultados.

El enfoque del análisis de sentimientos ha demostrado su validez para evaluar el estado de ánimo social de los ciudadanos en diferentes escalas temporales, registrando los distintos clústeres que surgieron, calibrando los estados de ánimo públicos a través de la valencia colectiva y detectando la prevalencia de las distintas emociones en las sucesivas fases de la pandemia.

El enfoque también ha mostrado, aunque de forma algo indirecta, el apoyo social a las políticas públicas. La combinación en modelos formales de información objetiva y subjetiva, en este caso sobre el proceso de vacunación del Covid-19 en España, puede proporcionar una visión más precisa de la realidad social, lo que permitiría una resolución más eficaz de los problemas.

En una línea similar, la erupción de Cumbre Vieja ("Tajogaite") se muestra en este trabajo como un valioso ejemplo de cómo la aplicación de enfoques novedosos y la combinación de herramientas pueden proporcionar nuevos marcos para entender los sentimientos y necesidades de la gente en el contexto de diversos tipos de riesgos de desastre.

Los resultados aquí obtenidos permiten concluir que el análisis de sentimientos aplicado a medios de prensa es una técnica útil que, en combinación con métodos estadísticos multivariantes y clasificadores de aprendizaje automático, es capaz de clasificar un artículo determinado dentro en una serie de grupos o clusters y establecer correlaciones útiles. De hecho, se observa cómo las emociones y sentimientos expresados en un texto cambian con el tiempo en respuesta a la dirección de los acontecimientos y al impacto social generado.

Este enfoque permitiría mejorar el diseño de herramientas para estudiar la evolución de los impactos sociales y medioambientales de un fenómeno natural y evaluar cómo cambian con el tiempo. Cuando una población se ve afectada por una crisis o una catástrofe natural, es importante analizar las tres capas principales de comunicación: la inmediatez y la espontaneidad (medios sociales), las noticias de prensa en periódicos y en línea (medios de prensa), y las reflexiones e informes en profundidad (artículos y reseñas). Estas tres modalidades de comunicación tienen su propia función social durante una crisis o catástrofe, con sus respectivas características, ventajas e inconvenientes.

En suma, los resultados de este estudio pueden contribuir a colmar lagunas de conocimiento relacionadas con posibles sesgos en la difusión de información durante crisis y catástrofes naturales, y a mejorar el análisis de las estrategias seguidas para reducir el riesgo y las consecuencias sociales de catástrofes.

Finalmente, la propia inteligencia artificial está contribuyendo a cuestionar asunciones tradicionales sobre las dinámicas emocionales y a evidenciar, a través de las investigaciones en *sentiment análisis*, las nuevas posibilidades de estudios mucho más amplios sobre los fenómenos sociales de nuestro tiempo. Constituye realmente un nuevo paradigma que se está difundiendo en múltiples ramas de las ciencias sociales, la economía, las ciencias políticas, y las ciencias de la comunicación. Y, por supuesto, en las propias ciencias y técnicas de la computación.

9. BIBLIOGRAFÍA

9. BIBLIOGRAFÍA

Agarwal, A., Singh, R., Toshniwal, D., 2018. Geospatial sentiment analysis using twitter data for UK-EU referendum. Journal of Information and Optimization Sciences 39, 303–317. https://doi.org/10.1080/02522667.2017.1374735

Barrett, L.F., 2006. Solving the Emotion Paradox: Categorization and the Experience of Emotion. Personality and Social Psychology Review 10, 20–46. https://doi.org/10.1207/s15327957pspr1001_2

Bernal, A.I., 2013. Audiencias y temas de noticias en medios online y Twitter. El caso de elpais.com, in: V Congreso Internacional de Ciberperiodismo y Web 2.0. Audiencias Activas y Periodismo.

Bohlouli, M., Dalter, J., Dornhöfer, M., Zenkert, J., Fathi, M., 2015. Knowledge discovery from social media using big data-provided sentiment analysis (SoMABiT). J Inf Sci 41, 779–798. https://doi.org/10.1177/0165551515602846

Booth, R., 2000. Loneliness as a component of psychiatric disorders. Medscape General Medicine 2, 1–7.

Bose, R., Dey, R.K., Roy, S., Sarddar, D., 2020. Sentiment Analysis on Online Product Reviews. pp. 559–569. https://doi.org/10.1007/978-981-13-7166-0_56

Budiharto, W., Meiliana, M., 2018. Prediction and analysis of Indonesia Presidential election from Twitter using sentiment analysis. J Big Data 5, 51. https://doi.org/10.1186/s40537-018-0164-1

Cacioppo, J., Cacioppo, S., 2014. Loneliness is a modern epidemic in need of treatment. https://www. newscientist. com/article/dn26739-loneliness-is-a-modern-epidemic-in-need-of-treatment/) 19, 2016.

Caetano, J.A., Lima, H.S., Santos, M.F., Marques-Neto, H.T., 2018. Using sentiment analysis to define twitter political users' classes and their homophily during the 2016 American presidential election. Journal of Internet Services and Applications 9, 18. https://doi.org/10.1186/s13174-018-0089-0

Capuano, N., Greco, L., Ritrovato, P., Vento, M., 2021. Sentiment analysis for customer relationship management: an incremental learning approach. Applied Intelligence 51, 3339–3352. https://doi.org/10.1007/s10489-020-01984-x

Carvalho, J., Plastino, A., 2021. On the evaluation and combination of state-of-the-art features in Twitter sentiment analysis. Artif Intell Rev 54, 1887–1936. https://doi.org/10.1007/s10462-020-09895-6

Chan, C.C., Liszka, K.J., 2013. Application of Rough Set Theory to Sentiment Analysis of Microblog Data. Intelligent Systems Reference Library 43. https://doi.org/10.1007/978-3-642-30341-8_10

Clavel, C., Callejas, Z., 2016. Sentiment Analysis: From Opinion Mining to Human-Agent Interaction. IEEE Trans Affect Comput. https://doi.org/10.1109/TAFFC.2015.2444846

Clynes, M., 1988. Generalised Emotion How it may be produced, and Sentic Cycle Therapy, in: Emotions and Psychopathology. Springer US, Boston, MA, pp. 107–170. https://doi.org/10.1007/978-1-4757-1987-1_6

Collins, K.P., 1991. On the Automation of Knowledge within Central Nervous Systems, in: Manuscript Presented at the AAAS Annual Meeting, Washington DC.

Collins, K.P., Marijuán, P.C., 1997. El Cerebro Dual: Un Acercamiento Interdisciplinar a la Naturaleza del Conocimiento Humano y Biológico. Editorial Hacer, Barcelona.

Csany, V., 1988. Contribution of the Genetical and Neural Memory to Animal Intelligence, in: Jerison, H., Jerison, I. (Eds.), Intelligence and Evolutionary Biology. Springer-Verlag, Berlin, pp. 299–318.

Dehaene, S., 2009. Reading in the brain. Penguin, New York.

Ebadi, A., Xi, P., Tremblay, S., Spencer, B., Pall, R., Wong, A., 2021. Understanding the temporal evolution of COVID-19 research through machine learning and natural language processing. Scientometrics 126, 725–739. https://doi.org/10.1007/s11192-020-03744-7

Ekman, P., 1999. Handbook of Cognition and Emotion, in: Dalgleish, T., Power, M. (Eds.), . John Wiley & Sons, Sussex.

Ekman, P., 1992. An argument for basic emotions. Cogn Emot 6, 169–200. https://doi.org/10.1080/02699939208411068

Garcia, K., Berton, L., 2021. Topic detection and sentiment analysis in Twitter content related to COVID-19 from Brazil and the USA. Appl Soft Comput 101, 107057. https://doi.org/10.1016/j.asoc.2020.107057

Granovetter, M., 2018. The impact of social structure on economic outcomes. The Sociology of Economic Life, Third Edition. https://doi.org/10.4324/9780429494338

Granovetter, M.S., 1973. The Strength of Weak Ties. American Journal of Sociology 78, 1360–1380. https://doi.org/10.1086/225469

Grant, N., Hamer, M., Steptoe, A., 2009. Social isolation and stress-related cardiovascular, lipid, and cortisol responses. Annals of Behavioral Medicine 37. https://doi.org/10.1007/s12160-009-9081-z

H. Manguri, K., N. Ramadhan, R., R. Mohammed Amin, P., 2020. Twitter Sentiment Analysis on Worldwide COVID-19 Outbreaks. Kurdistan Journal of Applied Research 54–65. https://doi.org/10.24017/covid.8

Han, T., Liu, C., Yang, W., Jiang, D., 2019. A novel adversarial learning framework in deep convolutional neural network for intelligent diagnosis of mechanical faults. Knowl Based Syst 165, 474–487. https://doi.org/10.1016/j.knosys.2018.12.019

Ilyas, S.H.W., Soomro, Z.T., Anwar, A., Shahzad, H., Yaqub, U., 2020. Analyzing Brexit's impact using sentiment analysis and topic modeling on Twitter discussion, in: The 21st Annual International Conference on Digital Government Research. ACM, New York, NY, USA, pp. 1–6. https://doi.org/10.1145/3396956.3396973

Jain, P.K., Pamula, R., Srivastava, G., 2021a. A systematic literature review on machine learning applications for consumer sentiment analysis using online reviews. Comput Sci Rev 41, 100413. https://doi.org/10.1016/j.cosrev.2021.100413

Jain, P.K., Saravanan, V., Pamula, R., 2021b. A Hybrid CNN-LSTM: A Deep Learning Approach for Consumer Sentiment Analysis Using Qualitative User-Generated Contents. ACM Transactions on Asian and Low-Resource Language Information Processing 20, 1–15. https://doi.org/10.1145/3457206

Jiménez-Zafra, S.M., Martín-Valdivia, M.T., Molina-González, M.D., Ureña-López, L.A., 2019. How do we talk about doctors and drugs? Sentiment analysis in forums expressing opinions for medical domain. Artif Intell Med 93, 50–57. https://doi.org/10.1016/j.artmed.2018.03.007

Kang, D., Park, Y., 2014. Review-based measurement of customer satisfaction in mobile service: Sentiment analysis and VIKOR approach. Expert Syst Appl 41, 1041–1050. https://doi.org/10.1016/j.eswa.2013.07.101

Kort, B., Reilly, R., Picard, R.W., 2001. An affective model of interplay between emotions and learning: Reengineering educational pedagogy-building a learning companion, in: Proceedings - IEEE International Conference on Advanced Learning Technologies, ICALT 2001. pp. 43–46. https://doi.org/10.1109/ICALT.2001.943850

Kumar, S., Yadava, M., Roy, P.P., 2019. Fusion of EEG response and sentiment analysis of products review to predict customer satisfaction. Information Fusion 52, 41–52. https://doi.org/10.1016/j.inffus.2018.11.001

Larg, A., Moss, J.R., 2011. Cost-of-illness studies: A guide to critical evaluation. Pharmacoeconomics 29. https://doi.org/10.2165/11588380-000000000-00000

Marijuán, P.C., Montero-Marín, J., Navarro, J., García-Campayo, J., del Moral, R., 2017. The "sociotype" construct: Gauging the structure and dynamics of human sociality. PLoS One 12, e0189568. https://doi.org/10.1371/journal.pone.0189568

Marijuán, P.C., Navarro, J., 2020. Sociotype and cultural evolution: The acceleration of cultural change alongside industrial revolutions. Biosystems 195, 104170. https://doi.org/10.1016/j.biosystems.2020.104170

Marijuan, P.C., Navarro, J., 2010. The Bonds of Laughter: A Multidisciplinary Inquiry into the Information Processes of Human Laughter. http://arxiv.org/abs/1010.5602.

McPherson, M., Smith-Lovin, L., Brashears, M.E., 2006. Social isolation in America: Changes in core discussion networks over two decades. Am Sociol Rev 71. https://doi.org/10.1177/000312240607100301

Mendon, S., Dutta, P., Behl, A., Lessmann, S., 2021. A Hybrid Approach of Machine Learning and Lexicons to Sentiment Analysis: Enhanced Insights from Twitter Data of Natural Disasters. Information Systems Frontiers 23, 1145–1168. https://doi.org/10.1007/s10796-021-10107-x

Mihalopoulos, C., Le, L.K.D., Chatterton, M. Lou, Bucholc, J., Holt-Lunstad, J., Lim, M.H., Engel, L., 2020. The economic costs of loneliness: a review of cost-of-illness and economic evaluation studies. Soc Psychiatry Psychiatr Epidemiol. https://doi.org/10.1007/s00127-019-01733-7

Mohammad, S.M., 2021. Sentiment analysis, in: Emotion Measurement. Elsevier, pp. 323–379. https://doi.org/10.1016/B978-0-12-821124-3.00011-9

Navarro, J., Cañete, M., Olivera, F.J., Gil-Lacruz, M., Gil-Lacruz, A., Marijuán, P.C., 2022. The Cost of Loneliness: Assessing the Social Relationships of the Elderly via an Abbreviated Sociotype Questionnaire for inside and outside the Clinic. Int J Environ Res Public Health 19, 1253. https://doi.org/10.3390/ijerph19031253

Navarro, J., del Moral, R., Alonso, M.F., Loste, P., Garcia-Campayo, J., Lahoz-Beltra, R., Marijuán, P.C., 2014. Validation of laughter for diagnosis and evaluation of depression. J Affect Disord 160, 43–49. https://doi.org/10.1016/j.jad.2014.02.035

Navarro, J., del Moral, R., Cuesta-Alvaro, P., Lahoz-Beltra, R., Marijuán, P.C., 2016. The entropy of laughter: discriminative power of laughter´s entropy in the diagnosis of depression. Entropy 18, 36.

Navarro, J., Marijuán, P.C., 2023. Natural intelligence and the 'economy' of social emotions. Biosystems 105039. https://doi.org/10.1016/j.biosystems.2023.105039

Navarro, J., Marijuán, P.C., 2022. The Natural, Artificial, and Social Domains of Intelligence: A Triune Approach, in: IS4SI 2021. MDPI, Basel Switzerland, p. 2. https://doi.org/10.3390/proceedings2022081002

Navarro, J., Piña, J.U., Mas, F.M., Lahoz-Beltra, R., 2023. Press media impact of the Cumbre Vieja volcano activity in the island of La Palma (Canary Islands): A machine learning and sentiment analysis of the news published during the volcanic eruption of 2021. International Journal of Disaster Risk Reduction 91, 103694. https://doi.org/10.1016/j.ijdrr.2023.103694

Neppalli, V.K., Caragea, C., Squicciarini, A., Tapia, A., Stehle, S., 2017. Sentiment analysis during Hurricane Sandy in emergency response. International Journal of Disaster Risk Reduction 21, 213–222. https://doi.org/10.1016/j.ijdrr.2016.12.011

Öztürk, N., Ayvaz, S., 2018. Sentiment analysis on Twitter: A text mining approach to the Syrian refugee crisis. Telematics and Informatics 35, 136–147. https://doi.org/10.1016/j.tele.2017.10.006

Pankaj, Pandey, P., Muskan, Soni, N., 2019. Sentiment Analysis on Customer Feedback Data: Amazon Product Reviews, in: 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon). IEEE, pp. 320–322. https://doi.org/10.1109/COMITCon.2019.8862258

Parrott, W.G., 2001. Emotions in Social Psychology: Essential Readings, Ann Arbor: Edwards.

Patel, R., Passi, K., 2020. Sentiment Analysis on Twitter Data of World Cup Soccer Tournament Using Machine Learning. IoT 1, 218–239. https://doi.org/10.3390/iot1020014

Păvăloaia, V.-D., Teodor, E.-M., Fotache, D., Danileţ, M., 2019. Opinion Mining on Social Media Data: Sentiment Analysis of User Preferences. Sustainability 11, 4459. https://doi.org/10.3390/su11164459

Pentland, A., 2014. Social Physics: How Good Ideas Spread-The Lessons from a New Science. The Penguin Press, New York, NY.

Picard, R., 1997. Affective Computing. MIT Press, Cambridge, MA.

Pinker, S., 2009. How the Mind Works. W. W. Norton, New York, NY.

Plutchik, R., 1980. A General Psychoevolutionary Theory of Emotion, in: Theories of Emotion. Elsevier, pp. 3–33. https://doi.org/10.1016/B978-0-12-558701-3.50007-7

Posner, J., Russell, J.A., Peterson, B.S., 2005. The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. Dev Psychopathol 17, 715–734. https://doi.org/10.1017/S0954579405050340

Putnam, R., 2000. Bowling alone. Simon & Schuster.

Qualter, P., Vanhalst, J., Harris, R., Roekel, E. Van, Lodder, G., Bangee, M., Maes, M., Verhagen, M., 2015. Loneliness Across the Life Span. Perspectives on Psychological Science 10. https://doi.org/10.1177/1745691615568999

Roccetti, M., Marfia, G., Salomoni, P., Prandi, C., Zagari, R.M., Gningaye Kengni, F.L., Bazzoli, F., Montagnani, M., 2017. Attitudes of Crohn's Disease Patients: Infodemiology Case Study and Sentiment Analysis of Facebook and Twitter Posts. JMIR Public Health Surveill 3, e51. https://doi.org/10.2196/publichealth.7004

Russell, J.A., 1980. A circumplex model of affect. J Pers Soc Psychol 39 (6), 1161–1178. https://doi.org/10.1037/h0077714

Russell, J.A., Barrett, L.F., 1999. Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant. J Pers Soc Psychol 76, 805–819. https://doi.org/10.1037/0022-3514.76.5.805

Salazar, J.L., Piles, J.J., Ruíz, J., Moreno-Jiménez, J.M., 2008. E-cognocracy and its voting process. Comput Stand Interfaces 30, 124–131. https://doi.org/10.1016/j.csi.2007.08.017

Salazar, J.L., Piles, J.J., Ruiz-Mas, J., Moreno-Jiménez, J.M., 2010. Security approaches in e-cognocracy. Comput Stand Interfaces 32, 256–265. https://doi.org/10.1016/j.csi.2010.01.004

Shankar, A., McMunn, A., Banks, J., Steptoe, A., 2011. Loneliness, Social Isolation, and Behavioral and Biological Health Indicators in Older Adults. Health Psychology 30. https://doi.org/10.1037/a0022826

Sharma, A., Ghose, U., 2020. Sentimental Analysis of Twitter Data with respect to General Elections in India. Procedia Comput Sci 173, 325–334. https://doi.org/10.1016/j.procs.2020.06.038

Shor, E., Roelfs, D.J., Yogev, T., 2013. The strength of family ties: A meta-analysis and meta-regression of self-reported social support and mortality. Soc Networks 35. https://doi.org/10.1016/j.socnet.2013.08.004

Tang, J.E., Arvind, V., White, C.A., Dominy, C., Kim, J.S., Cho, S.K., Walsh, A., 2023. Using Sentiment Analysis to Understand What Patients Are Saying About Hand Surgeons Online. HAND 18, 854–860. https://doi.org/10.1177/15589447211060439

Tomasello, M., 2019. Becoming human. Harvard University Press, Cambridge.

Toncovich, A., Turón, A., Escobar, M.T., Moreno-Jiménez, J.M., 2013. A Quantitative Approach to Identify the Arguments that Support Decisions in E-Cognocracy, in: Governance, Communication, and Innovation in a Knowledge Intensive Society. IGI Global, pp. 64–76. https://doi.org/10.4018/978-1-4666-4157-0.ch006

Trivers, R., 1985. Social Evolution. Benjamin-Cummings Publishing Co, San Francisco.

Turón, A., Altuzarra, A., Moreno-Jiménez, J.M., Navarro, J., 2023. Evolution of social mood in Spain throughout the COVID-19 vaccination process: a machine learning approach to tweets analysis. Public Health 215, 83–90. https://doi.org/10.1016/j.puhe.2022.12.003

van Kleef, G.A., Fischer, A.H., 2016. Emotional collectives: How groups shape emotions and emotions shape groups. Cogn Emot 30, 3–19. https://doi.org/10.1080/02699931.2015.1081349

Vinodhini, G., Chandrasekaran, R., 2012. Sentiment Analysis and Opinion Mining : A Survey International Journal of Advanced Research in Sentiment Analysis and Opinion Mining : A Survey. International Journal of Advanced Research in Computer Science and Software Engineering 2.

Wankhade, M., Rao, A.C.S., Kulkarni, C., 2022. A survey on sentiment analysis methods, applications, and challenges. Artif Intell Rev 55, 5731–5780. https://doi.org/10.1007/s10462-022-10144-1

Watt Smith, T., 2021. The Book of Human Emotions: An Encyclopedia of Feeling from Anger to Wanderlust [WWW Document]. https://web.archive.org/web/20210418032350/http://anarchiveforemotions.com/files/DisOrder_uploads/images/TheBookOfHumanEmotions.pdf.

Weissman, G.E., Ungar, L.H., Harhay, M.O., Courtright, K.R., Halpern, S.D., 2019. Construct validity of six sentiment analysis methods in the text of encounter notes of patients with critical illness. J Biomed Inform 89, 114–121. https://doi.org/10.1016/j.jbi.2018.12.001

WHO, 2020. Mental Health and Psychosocial Considerations During COVID-19 Outbreak. World Health Organization.

Xing, F.Z., Cambria, E., Welsch, R.E., 2018. Natural language based financial forecasting: a survey. Artif Intell Rev 50, 49–73. https://doi.org/10.1007/s10462-017-9588-9

Yuan, F., Li, M., Liu, R., 2020. Understanding the evolutions of public responses using social media: Hurricane Matthew case study. International Journal of Disaster Risk Reduction 51, 101798. https://doi.org/10.1016/j.ijdrr.2020.101798

Zuboff, S., 2019. Surveillance Capitalism and the Challenge of Collective Action. New Labor Forum 28, 10–29. https://doi.org/10.1177/1095796018819461

APÉNDICE I.

INFORMACIÓN ÚTIL DE LAS PUBLICACIONES QUE SE RECOGEN EN LA TESIS

APÉNDICE I. INFORMACIÓN ÚTIL DE LAS PUBLICACIONES QUE SE RECOGEN EN LA TESIS

- Navarro, J.*, Cañete, M., Olivera, F. J., Gil-Lacruz, M., Gil-Lacruz, A., & Marijuán, P. C. (2022). The Cost of Loneliness: Assessing the Social Relationships of the Elderly via an Abbreviated Sociotype Questionnaire for inside and outside the Clinic. *International Journal of Environmental Research and Public Health*, 19(3), 1253.
https://doi.org/10.3390/ijerph19031253

    - CATEGORÍA SJR: PUBLIC, ENVIRONMENTAL & OCCUPATIONAL HEALTH
    - FACTOR DE IMPACTO SJR 2022: 0,828
    - CUARTIL: 2º

    CONTRIBUCIÓN PERSONAL:

    - Contribución al diseño del procedimiento experimental.
    - Tramitación de la aprobación por el Comité Ético de Investigación Clínica de Aragón (CEICA).
    - Contribución al análisis estadístico de los datos.
    - Contribución a la discusión de los resultados obtenidos y a la preparación del manuscrito.
    - Envío del articulo a la revista como Corresponding Author.
    - Discusión de las revisiones recibidas y preparación de la versión definitiva.

- Navarro, J.* & Marijuán, P. C. (2023). Natural intelligence and the ‘economy’ of social emotions: A connection with AI sentiment analysis. *Biosystems*, 233, 105039.
https://doi.org/10.1016/j.biosystems.2023.105039

    - CATEGORÍA JCR SCIE: MATHEMATICAL & COMPUTATIONAL BIOLOGY
    - FACTOR DE IMPACTO JCR 2022: 1,6
    - CUARTIL: 3º

    CONTRIBUCIÓN PERSONAL:

    - Contribución a la hipótesis inicial de la "economía" global de las emociones sociales.
    - Contribución al diseño del estudio, revisión y obtención bibliográfica.
    - Contribución a la escritura, discusión de los resultados obtenidos y a la preparación del manuscrito.
    - Realización personal de las figuras 1 y 2.
    - Envío del articulo a la revista como Corresponding Author.
    - Discusión de las revisiones recibidas y preparación de la versión definitiva.

- Turón, A., Altuzarra, A., Moreno-Jiménez, J. M., & Navarro, J.* (2023). Evolution of social mood in Spain throughout the COVID-19 vaccination process: a machine learning approach to tweets analysis. *Public health*, 215, 83-90. https://doi.org/10.1016/j.puhe.2022.12.003

  - CATEGORÍA JCR SSCI: PUBLIC, ENVIRONMENTAL & OCCUPATIONAL HEALTH
  - FACTOR DE IMPACTO JCR 2022: 5,2
  - CUARTIL: 1º

  CONTRIBUCIÓN PERSONAL:

  - Contribución al diseño del procedimiento experimental (elección temática, fechas…).
  - Desarrollo de los algoritmos para la realización del análisis de sentimientos.
  - Desarrollo del análisis de sentimiento de los tweets.
  - Contribución al análisis estadístico de los datos.
  - Contribución a la discusión de los resultados obtenidos y a la preparación del manuscrito.
  - Envío del articulo a la revista como Corresponding Author y preparación de la versión definitiva.

- Navarro, J., Pina, J. U., Mas, F. M., & Lahoz-Beltra, R. (2023). Press media impact of the Cumbre Vieja volcano activity in the island of La Palma (Canary Islands): A machine learning and sentiment analysis of the news published during the volcanic eruption of 2021. *International Journal of Disaster Risk Reduction*, 91, 103694. https://doi.org/10.1016/j.ijdrr.2023.103694

  - CATEGORÍA JCR SCIE: GEOSCIENCES, MULTIDISCIPLINARY
  - FACTOR DE IMPACTO JCR 2022: 5,0
  - CUARTIL: 1º

  CONTRIBUCIÓN PERSONAL:

  - Contribución al diseño del procedimiento experimental (elección temática, prensa nacional e internacional…).
  - Desarrollo del análisis de sentimiento de las noticias.
  - Contribución al análisis estadístico y visualización gráfica de los datos.
  - Contribución a la discusión de los resultados obtenidos y a la preparación del manuscrito.