

DOI: <http://doi.org/10.22585/hospdomic.v7i4.200>

Topic Modeling mediante Machine Learning no supervisado de artículos científicos sobre Salud Laboral y Servicios de Atención de Salud a Domicilio

Topic Modeling through unsupervised Machine Learning of scientific articles on Occupational Health and Home Care Services

Rubén Palomo-Llinares¹  0000-0002-1890-4337

Julia Sánchez-Tormo²  0000-0001-9341-8737

1. Universidad Miguel Hernández, Departamento de Salud Pública e Historia de la Ciencia, Sant Joan d Alacant, Alicante, España.

2. Instituto de Investigación Sanitaria y Biomédica de Alicante (ISABIAL), Alicante, España.

Correspondencia/Correspondence

Rubén Palomo-Llinares
palomo.rub@gmail.com

Recibido/Received

03.10.2023

Aceptado/Accepted

29.10.2023

Conflicto de Intereses/Competing interest

No existe ningún conflicto de interés.

Financiación/Funding

Este trabajo no ha recibido ninguna financiación.

Contribuciones de autoría/Author contributions

RPL y JST han contribuido por igual a la autoría de este trabajo.

CÓMO CITAR ESTE TRABAJO | HOW TO CITE THIS PAPER

Palomo-Llinares R, Sánchez-Tormo J. Topic Modeling mediante Machine Learning no supervisado de artículos científicos sobre Salud Laboral y Servicios de Atención de Salud a Domicilio. *Hosp Domic.* 2023;7(4):167-78.

RESUMEN

Objetivo: Identificar de manera no supervisada mediante topic modeling los temas de mayor interés en el campo de la Salud Laboral y los Servicios de Atención a Domicilio de los artículos científicos publicados en la materia.

Método: Este estudio empleó el algoritmo de Machine Learning no supervisado Asignación Latente de Dirichlet para el topic modeling y el lexicon NRC para la realización del análisis de sentimientos del corpus de las fichas documentales obtenidas de MEDLINE (vía PubMed) usando los descriptores “Salud Laboral” y “Servicios de Atención de Salud a Domicilio”.

Resultados: Del total de 70 fichas documentales analizadas, se obtuvo que la intensidad de las emociones en los textos era baja (oscilando en valores de 5 a 10), teniendo una mayor representación los sentimientos positivos frente a los negativos en una relación de 60/40. No hubo una variación de las proporciones de las emociones con respecto al período del estudio. Se identificaron los cuatro temas de mayor interés en los artículos analizados: cuidado domiciliario y satisfacción de los cuidadores, período de lactancia, programas de rehabilitación, y actividad física para mitigación del dolor.

Conclusiones: Se ha podido constatar que las metodologías del procesado de lenguaje natural pueden ser una gran herramienta de apoyo al análisis de artículos científicos. Concretamente, se ha logrado determinar de manera clara y no supervisada los temas de mayor interés en el campo de la Salud Laboral y la Atención de Salud a Domicilio.

Palabras clave: Salud Laboral; Servicios de Atención de Salud a Domicilio; topic modeling; análisis de sentimiento.

ABSTRACT

Objective: To identify in an unsupervised manner through topic modeling the topics of greatest interest in the field of Occupational Health and Home Care Services from the scientific articles published on the subject.

Method: The study used the unsupervised Machine Learning algorithm Dirichlet Latent Assignment for topic modeling and the NRC lexicon to carry out the sentiment analysis of the corpus of document files obtained from MEDLINE (via PubMed) using the descriptors “Occupational Health” and “Home Care Services”.

Results: Of the total of 70 documentary files analyzed, it was obtained that the intensity of the emotions in the texts was low (ranging in values from 5 to 10), with positive feelings having a greater representation compared to negative ones in a ratio of 60/40. There was no variation in the proportions of emotions with respect to the study period. The four topics of greatest interest were identified in the articles analyzed: home care and caregiver satisfaction, breastfeeding period, rehabilitation programs, and physical activity to mitigate pain.

Conclusions: It has been confirmed that natural language processing methodologies can be a great support tool for the analysis of scientific articles. Specifically, it has been possible to determine in a clear and unsupervised manner the topics of greatest interest in the field of Occupational Health and Home Care Services.

Keywords: Occupational Health; Home Care Services; topic modeling; sentiment analysis.

INTRODUCCIÓN

En los últimos años, los modelos de obtención de insights basados en las técnicas de Machine Learning han ido obteniendo cada vez mayor reconocimiento en todos los ámbitos de las tecnologías de la información⁽¹⁾, desde el desarrollo de campañas de márketing personalizadas⁽²⁾, al uso extensivo de tomas de decisiones en sectores financieros^(3,4) como al campo del procesado del lenguaje natural⁽⁵⁾.

Dentro del campo del procesado del lenguaje natural, existen dos aplicaciones específicas que pueden ser especialmente convenientes para el campo de la salud pública. El primero de ellos es el análisis de sentimiento (AS)⁽⁶⁾, también conocido como minería de opinión, que se encarga de identificar y extraer información subjetiva de los textos analizados, de una manera objetiva y sistemática, minimizando los tiempos y cuantificando los resultados obtenidos. Por otra parte, tenemos el modelado de temas o topic modeling (TM)⁽⁷⁾, que permite extraer estructuras semánticas subyacentes a los textos analizados, que podrían ocasiones permanecer ocultas, de manera automatizada. Una de las principales ventajas de esta metodología es que permite realizar el análisis de los textos de una manera rápida, eficiente y sistematizada, aumentando su valor científico por su alta reproducibilidad y objetividad⁽⁸⁾.

En este sentido, es lógico pensar que las bondades de esta metodología se deberían de poder emplear con éxito en el campo de los estudios de la salud pública. De esta manera, algunos autores ya han empezado a emplear éstas metodologías para analizar textos como encuestas de satisfacción de pacientes en el ámbito de las consultas sanitarias^(9,10). Pero también se están empezando a emplear estas técnicas a la revisión de literatura científica específica, tanto para la obtención de insights específicos⁽¹¹⁻¹³⁾, herramientas de análisis de texto en bases de datos⁽¹⁴⁾, como herramienta alternativa a las metodologías clásicas de revisiones sistemáticas manuales⁽⁸⁾.

Además, el empleo de AS también ha sido empleado en las ciencias de la salud para diversos objetivos. Estudios de tendencias de violencia sexual en institutos mediante el uso de Twitter⁽¹⁵⁾, estudios epidemiológicos de las enfermedades transmitidas por los mosquitos⁽¹⁶⁾, estudio de riesgo de suicidios en la población⁽¹⁷⁾, herramienta de análisis de ciencias alimentarias y nutrición⁽¹⁸⁾, y por supuesto, también se ha empleado con éxito para la monitorización de brotes epidemiológicos durante la crisis del COVID-19⁽¹⁹⁾.

Por otra parte, debido a la inversión de la pirámide poblacional en los países occidentales⁽²⁰⁾, y a la mayor tendencia a la hospitalización a domicilio^(21,22), tanto para el cuidado de mayores como para el cuidado de enfermos paliativos, los Servicios de Atención de Salud a Domicilio se han ido viendo incrementados en los últimos años⁽²³⁻²⁵⁾. Es por todo lo expuesto que aumentar el conocimiento respecto a los puntos de interés y las necesidades reales de este sector es de vital importancia de cara a tener una buena previsión de recursos y medios por parte de las autoridades competentes en el ámbito de la Salud Pública.

Por todo ello, el objetivo de este estudio fue el de obtener, de manera automatizada y objetivable, mediante modelos no supervisados de Machine Learning, los temas de mayor interés en los artículos científicos que trataban de la Salud Pública y los Servicios de Atención de Salud a Domicilio.

MÉTODOS DISEÑO

Estudio descriptivo transversal mediante AS y TM de las fichas documentales recuperadas mediante técnica sistemática por su referencia.

Fuente de obtención de los datos

Los datos se obtuvieron por consulta directa y acceso, vía Internet, a la base de datos bibliográfica de ámbito científico-sanitario MEDLINE (vía PubMed).

Tratamiento de la información

La definición de los términos de búsqueda se realizó mediante consulta al *Thesaurus* de los Descriptores en Ciencias de la Salud (DeCS) desarrollado por el Centro Latinoamericano y del Caribe de Información en Ciencias Médicas (BIREME) junto con la equivalencia establecida por la U.S. National Library of Medicine, los Medical Subject Headings (MeSH).

Analizando ambos *Thesaurus* y sus fichas asociadas de indización (Entry Terms), se consideran las siguientes ecuaciones de búsqueda:

Ecuación 1: Salud Laboral (Occupational Health)

"Occupational Health"[Mesh] OR "Occupational Health"[Title/Abstract] OR "Industrial Hygiene"[Title/Abstract] OR "Industrial Health"[Title/Abstract] OR "Occupational Safety"[Title/Abstract] OR "Employee Health"[Title/Abstract] OR "Occupational Exposure"[Mesh] OR "Occupational Exposure"[Title/Abstract] OR "Occupational Stress"[Mesh] OR "Occupational Stress"[Title/Abstract] OR "Occupational Diseases"[Mesh] OR "Occupational Diseases"[Title/Abstract] OR "Occupational Hazard"[Title/Abstract] OR "Occupational Medicine"[Mesh] OR "Occupational Medicine"[Title/Abstract] OR "Occupational Health Safety"[Title/Abstract] OR "Occupational Stressor"[Title/Abstract] OR "Occupational Factor"[Title/Abstract] OR "Professional Health"[Title/Abstract] OR "Workplace"[Mesh] OR "Workplace"[Title/Abstract] OR "Workplace Health"[Title/Abstract] OR "Workplace Safety"[Title/Abstract] OR "Safety Climate"[Title/Abstract] OR "Total Worker Health"[Title/Abstract] OR "Working Environment"[Title/Abstract] OR "Job Satisfaction"[Mesh] OR "Job Satisfaction"[Title/Abstract] OR "Job Stress"[Title/Abstract] OR "Job Security"[Title/Abstract] OR "Psychosocial Working Conditions"[Title/Abstract] OR "Employment"[Mesh] OR "Employment"[Title/Abstract] OR "Labor Force"[Title/Abstract] OR "Labour Force"[Title/Abstract] OR "Precarious Employment"[Title/Abstract] OR "Marginal Employment"[Title/Abstract] OR "Employment Insecurity" OR "Employment Insecurities"[Title/Abstract] OR "Employment Status"[Title/Abstract] OR "Occupational Status"[Title/Abstract] OR "Underemployment"[Title/Abstract] OR "Employee Health"[Title/Abstract]

Ecuación 2: Servicios de Atención de Salud a Domicilio (Home Care Services)

"Home Care Services"[MeSH] OR "Home Care"[Title/Abstract] OR "Domiciliary Care"[Title/Abstract] OR "Home Care Services, Hospital- Based"[Mesh] OR "Hospital-Based Home Care"[All Fields] OR "Hospital Based Home Care"[Title/Abstract] OR "Hospital Home Care Services"[Title/Abstract] OR "Hospital-Based Home Care Services"[Title/Abstract] OR "Hospital Based Home Care Services"[Title/Abstract] OR "Home Hospitalization"[Title/Abstract] OR "Hospital at Home"[Title/Abstract]

tract] OR "Hospital-at-Home"[Title/Abstract] OR "Hospital Home Care"[Title/Abstract] OR "Hospital at Home Care"[Title/Abstract] OR "Hospital in the Home"[Title/Abstract]

La ecuación de búsqueda final se empleó en la base de datos MEDLINE, vía PubMed, mediante la unión booleana de las 2 ecuaciones: Ecuación 1 AND Ecuación 2. Además, se empleó el filtro de "Clinical Trials" para obtener la máxima evidencia científica.

La búsqueda se realizó el 4 de septiembre de 2023.

Selección final de los artículos

Los artículos que fueron escogidos para el AS y TM cumplieron los siguientes criterios:

- Inclusión: ser ensayo clínico, estar publicado en revistas arbitradas por pares y estar redactados en inglés.
- Exclusión: no se estableció ningún criterio de exclusión.

Solamente se incluyeron artículos en lengua inglesa para poder hacer uso del diccionario de emociones de la National Research Council de Canada (NRC lexicon)⁽²⁶⁾ por su extendido uso en este campo de investigación.

La selección de artículos pertinentes fue revisada por los autores del presente texto.

Extracción de los datos

El control de la corrección de los datos se realizó mediante el uso de dobles tablas. Esto permitió la posible detección de desviaciones y su posterior subsanación mediante nueva consulta a la fuente original de los datos.

Dado que todos los artículos fueron obtenidos de una única base de datos, no se debió tener en cuenta la gestión de artículos duplicados.

El almacenamiento de las fichas documentales de los artículos se almacenó como texto plano previo a su incorporación al sistema de análisis.

Análisis de los datos

Para el AS se utilizó el Diccionario NRC, que asocia cada palabra del corpus de artículos a una de las ocho emociones básicas definidas por Plutchik (ira, miedo, anticipación, confianza, sorpresa, tristeza, alegría y disgusto) y a dos sentimientos generales (negativo y positivo)⁽²⁷⁾. Esta asociación se empleó para la realización de la nube de palabras, del gráfico de radar y del gráfico de evolución temporal.

La relación entre los diferentes sentimientos se cuantificó mediante la prueba U de Mann-Witney, y la relación entre las diferentes emociones se cuantificó mediante la prueba de Kruskal-Wallis.

Para la realización del TM se empleó el algoritmo de Machine Learning no supervisado de la Asignación Latente de Dirichlet (ADL). Este algoritmo de clusterización de patrones realiza la suposición de que existen diferentes capas de patrones dentro de cada elemento de la población (no en número fijo). Además, asume que las distribuciones de las categorías siguen una distribución de Dirichlet⁽²⁸⁾.

El análisis de datos se realizó mediante el software R v4.2.2 con el paquete de trabajo RStudio 2023.09.0 build 463. La librería específica relevante al TM fue "topicmodels" v0.2-14, mientras que las librerías empleadas para el AS fueron la "tidytext" v0.4.1, la "ggadar" v0.2, la "ggstream" v0.1.0 y la "wordcloud" v2.6.

Aspectos éticos

Todos los datos fueron obtenidos de las fichas documentales que aparen libremente en la base de datos de MEDLINE. Por esto, y conforme a lo establecido por la Ley 14/2007, de investigación biomédica⁽²⁹⁾, no fue necesaria la aprobación del Comité de Ética e Investigación al utilizar datos secundarios.

RESULTADOS

Al aplicar los criterios de búsqueda se recuperaron un total de 1889 referencias, todas ellas de la base de datos de MEDLINE (vía PubMed). Ninguno de estas referencias tuvo que ser eliminado por motivos de duplicidades ni por otros motivos de calidad.

Tras depurar aquellos registros que no eran ensayos clínicos (1.819), el número de artículos que se consideraron para su procesamiento en el análisis fue de 70. Ver figura 1.

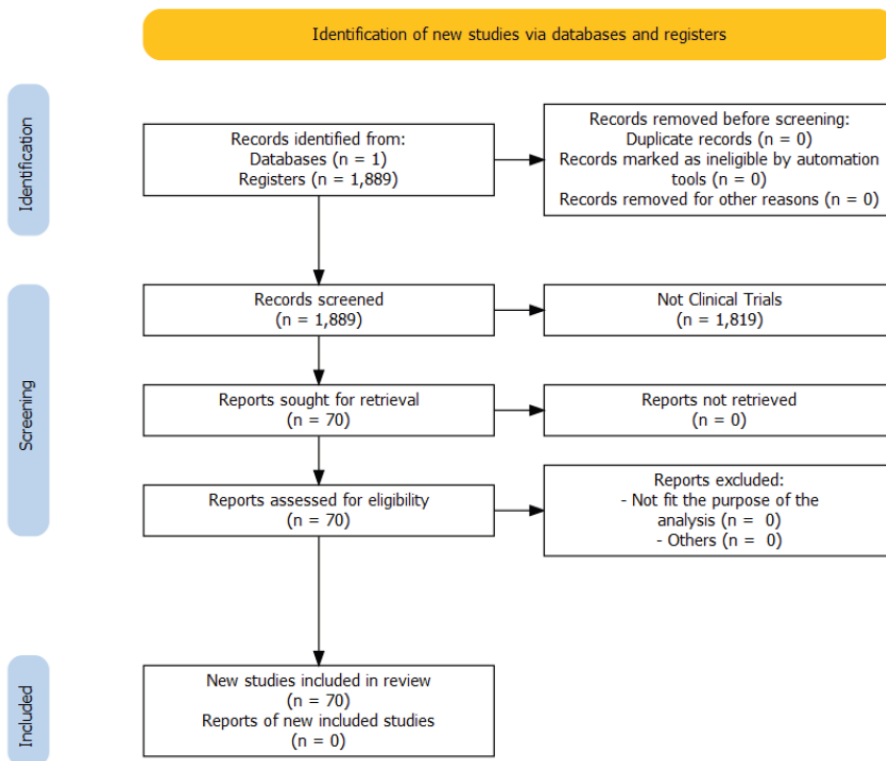


Figura 1: Identificación y selección de estudios. (Salida según registro del programa Rstudio).

En estas gráficas se cuantificaron que existían diferencias estadísticamente significativas entre los dos sentimientos, así como entre las diferentes emociones aparecidas en los artículos.

En el gráfico de radar se cuantificó la intensidad de aparición de cada una de las emociones de la rueda de Plutchik: ira, miedo, anticipación, confianza, sorpresa, tristeza, alegría y disgusto (ver figura 4). Como se puede apreciar en el gráfico, la intensidad de las emociones en general no es muy alta en ningún caso. La emoción que presentó una mayor intensidad fue confianza, y las que menos disgusto y sorpresa seguidas por alegría; quedando anticipación, miedo y tristeza en un rango medio de intensidad.

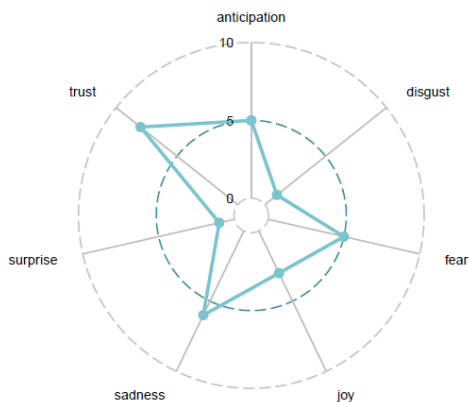


Figura 4: Gráfico de radar para la intensidad de los sentimientos según la rueda de Plutchik.

El gráfico de evolución temporal de las emociones (figura 5) mostró que prácticamente en todo el período de estudio, la relación de aparición de las emociones en los artículos científicos se mantuvo estable, presentando pocas variaciones. Tan solo se destacaría una ligera variación en los últimos dos años de un aumento de la emoción de anticipación a costa de la emoción de confianza.

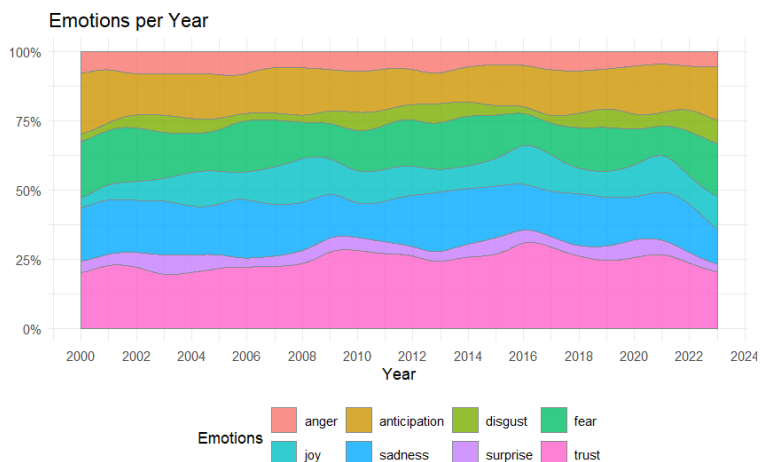


Figura 5: Evolución de las emociones de la rueda de Plutchik por año.

Topic modeling

En la figura 6 se puede apreciar el resultado obtenido de la aplicación del algoritmo de *Machine Learning* no supervisado ADL con 4 conjuntos de agrupaciones. Destacar que, en este tipo de metodologías, es el propio modelo el que se encarga de obtener las agrupaciones subyacentes (en este caso temas de los textos) a partir de la repetición de patrones y estructuras subyacentes en el corpus de textos analizados.

La Asignación Latente de Dirichlet trata cada documento como una mezcla de temas y cada tema como una mezcla de palabras, permitiendo que los documentos se superpongan por contenido en lugar de separarlos en grupos discretos, de manera que refleje el uso de lenguaje natural.

En el gráfico se muestran, para cada agrupación determinada por el algoritmo, aquellas palabras que tengan mayor influencia en los temas detectados. Esto se realiza cuantificando la frecuencia de aparición en las estructuras de los textos analizados.

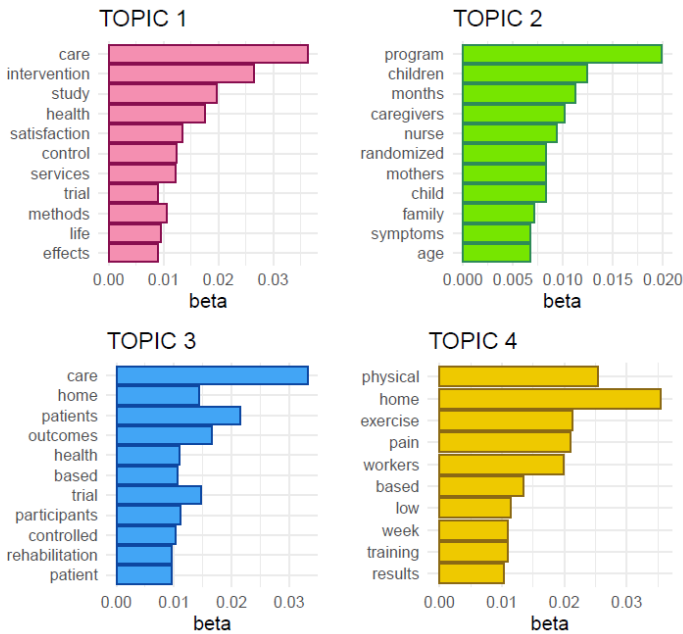


Figura 6: Resultados del TM para 4 agrupaciones.

Como se puede apreciar en el gráfico, los cuatro temas más predominantes que se encontraron en los textos fueron los siguientes:

- *TOPIC 1*: Estudios con un tema que impliquen una intervención general en el cuidado domiciliario y su relación con la satisfacción de los cuidadores.
- *TOPIC 2*: Estudios que tienen como foco el cuidado de los recién nacidos en los meses de lactancia de los trabajadores.
- *TOPIC 3*: Estudios que tengan como objetivo conocer el devenir de la salud de los trabajadores a partir de programas de rehabilitación en los domicilios.

- *TOPIC 4*: Estudios que tratan de la actividad física para mitigación del dolor de trabajadores.

DISCUSIÓN

En primer lugar, se puede apreciar que los términos más empleados en los artículos científicos se centran en el dolor, las intervenciones domiciliarias, el cáncer, el embarazo y el sobrepeso. Salvo el término “dolor”, que sí que tiene una mayor presencia, ningún otro término destaca por encima del resto.

Además, gracias al estudio de frecuencia de aparición de las emociones, se pudo cuantificar que sí que existía una ligera tendencia estadísticamente significativa tanto del sentimiento positivo frente al negativo, como de las emociones asociadas al espectro positivo frente aquellas del espectro negativo. Esto era algo esperable ya que, en el ámbito científico, se suele emplear un lenguaje neutro-positivo, pero sin grandes cargas emocionales⁽³⁰⁾.

Siguiendo con esta línea, en el gráfico de radar se puede corroborar esta idea, ya que se puede apreciar que la emoción con mayor intensidad estaba en el rango positivo de las emociones. Sin embargo, la intensidad de aparición de las emociones no es muy alta en ninguno de los ocho puntos cardinales del gráfico (oscilando en un rango de entre 10 y 5 puntos). El sentimiento de mayor prevalencia fue el de “confianza” y los que menos, “sorpresa” y “disgusto”. Como se ha comentado en el párrafo anterior, este resultado era el esperado debido a la general neutralidad empleada en la redacción de textos científicos⁽³⁰⁾.

En la evolución temporal de las emociones, podemos ver que las proporciones aparecidas en el gráfico se mantienen bastante constantes a lo largo de todo el período de estudio, no habiendo ningún evento temporal clave que cambie esta tendencia, ni siquiera la aparición del COVID-19.

Finalmente, respecto al modelado no supervisado, se han podido determinar los cuatro temas de mayor interés tratado en los artículos científicos dentro del campo estudiado. Estos fueron el tema de los cuidados domiciliarios, con el nivel de satisfacción que presentaban los cuidadores; el tema de la gestión de los meses de lactancia por parte de las trabajadoras; el tema de los programas de rehabilitación a domicilio; y finalmente el tema de la actividad física para mitigar el dolor en trabajadores. Como se pudo apreciar, esta obtención de resultados se realizó de manera autónoma y repetible, ya que fue el algoritmo de Machine Learning el que se encargó de la determinación de los temas por la superposición de estructuras del lenguaje natural dentro del corpus de textos presentes en las fichas documentales de los artículos analizados.

CONCLUSIONES

A la vista de los resultados, queda claro que las metodologías del procesado del lenguaje natural, tanto el AS, como los algoritmos de TM, pueden emplearse para extraer información subyacente de los textos científicos del ámbito de las ciencias de la salud de una manera objetiva y estructurada. En nuestro caso, se ha obtenido el resultado deseado, que era el de conocer de una manera clara y mediante técnicas no supervisadas, que evitan la introducción de sesgos humanos, los temas más tratados dentro del campo de la Salud Laboral y los Servicios de Atención de Salud a Domicilio.

BIBLIOGRAFÍA

1. Zippel C, Bohnet-Joschko S. Rise of Clinical Studies in the Field of Machine Learning: A Review of Data Registered in ClinicalTrials.gov. *Int J Environ Res Public Health*. 2021;18(10):5072. DOI: 10.3390/ijerph18105072
2. Ngai EWT, Wu Y. Machine learning in marketing: A literature review, conceptual framework, and research agenda. *J Bus Res*. 2022;145:35-48. DOI: 10.1016/j.jbusres.2022.02.049
3. Buczynski W, Cuzzolin F, Sahakian B. A review of machine learning experiments in equity investment decision-making: why most published research findings do not live up to their promise in real life. *Int J Data Sci Anal*. 2021;11(3):221-42. DOI: 10.1007/s41060-021-00245-5
4. Buchanan BG, Wright D. The impact of machine learning on UK financial services. *Oxf Rev Econ Policy*. 2021;37(3):537-63. DOI: 10.1093/oxrep/grab016
5. Khanbhai M, Anyadi P, Symons J, Flott K, Darzi A, Mayer E. Applying natural language processing and machine learning techniques to patient experience feedback: a systematic review. *BMJ Health Care Inform*. 2021;28(1):e100262. DOI: 10.1136/bmjhci-2020-100262
6. Análisis de sentimiento. En: Wikipedia, la enciclopedia libre [Internet]. 24 de mayo de 2023 [citado 1 de octubre de 2023]. Recuperado: <http://bit.ly/46B9uro>
7. Topic model. En: Wikipedia [Internet]. 17 de septiembre de 2023 [citado 1 de octubre de 2023]. Recuperado: <https://bit.ly/48CXwiA>
8. Lutz AM, Lutz R. Topic Modeling as an evaluation basis in literature research - A proposal for a new literature review method for machine-assisted source evaluation using the example of anthropology. *Anthropol Anz Ber Uber Biol-Anthropol Lit*. 2023;80(2):119-34. DOI: 10.1127/anthranz/2023/1660
9. Alexander G, Bahja M, Butt GF. Automating Large-scale Health Care Service Feedback Analysis: Sentiment Analysis and Topic Modeling Study. *JMIR Med Inform*. 2022;10(4):e29385. DOI: 10.2196/29385
10. Báez P, Arancibia AP, Chaparro MI, Bucarey T, Núñez F, Dunstan J. Procesamiento de lenguaje natural para texto clínico en español: el caso de las listas de espera en Chile. *Rev Médica Clínica Las Condes*. 2022;33(6):576-82. DOI: 10.1016/j.rmclc.2022.10.002
11. Scarpino I, Zucco C, Vallengunga R, Luzza F, Cannataro M. Investigating Topic Modeling Techniques to Extract Meaningful Insights in Italian Long COVID Narration. *Biotech Basel Switz*. 2022;11(3):41. DOI: 10.3390/biotech11030041
12. Muchene L, Safari W. Two-stage topic modelling of scientific publications: A case study of University of Nairobi, Kenya. *PloS One*. 2021;16(1):e0243208. DOI: 10.1371/journal.pone.0243208
13. Kavvadias S, Drosatos G, Kaldoudi E. Supporting topic modeling and trends analysis in biomedical literature. *J Biomed Inform*. 2020;110:103574. DOI: 10.1016/j.jbi.2020.103574
14. Liu L, Tang L, Dong W, Yao S, Zhou W. An overview of topic modeling and its current applications in bioinformatics. *SpringerPlus*. 2016;5(1):1608. DOI: 10.1186/s40064-016-3252-8
15. Xue J, Zhang B, Zhang Q, Hu R, Jiang J, Liu N, et al. Using Twitter-Based Data for Sexual Violence Research: Scoping Review. *J Med Internet Res*. 2023;25:e46084. DOI: 10.2196/46084
16. Ong S-Q, Pauzi MBM, Gan KH. Text mining in mosquito-borne disease: A systematic review. *Acta Trop*. 2022;231:106447. DOI: 10.1016/j.actatropica.2022.106447

17. Castillo-Sánchez G, Marques G, Dorrnzoro E, Rivera-Romero O, Franco-Martín M, De la Torre-Díez I. Suicide Risk Assessment Using Machine Learning and Social Networks: a Scoping Review. *J Med Syst.* 2020;44(12):205. DOI: 10.1007/s10916-020-01669-5
18. Tao D, Yang P, Feng H. Utilization of text mining as a big data analysis tool for food science and nutrition. *Compr Rev Food Sci Food Saf.* 2020;19(2):875-94. DOI: 10.1111/1541-4337.12540
19. Stefanis C, Giorgi E, Kalentzis K, Tselemonis A, Nena E, Tsigalou C, et al. Sentiment analysis of epidemiological surveillance reports on COVID-19 in Greece using machine learning models. *Front Public Health.* 2023;11:1191730. DOI: 10.3389/fpubh.2023.1191730
20. O'Sullivan JN. Trends in population health and demography. *The Lancet.* Elsevier; 2021;398(10300):580. DOI: 10.1016/S0140-6736(21)01050-3
21. Hanley E. The role of home care in palliative care services. *Care Manag J J Case Manag J Long Term Home Health Care.* 2004;5(3):151-7. DOI: 10.1891/cmaj.2004.5.3.151
22. Seow H, Barbera L, Howell D, Dy SM. Using more end-of-life homecare services is associated with using fewer acute care services: a population-based cohort study. *Med Care.* 2010;48(2):118-24. DOI: 10.1097/MLR.0b013e3181c162ef
23. Elliott J, Gordon A, Tong CE, Stolee P. «We've got the home care data, what do we do with it?»: understanding data use in decision making and quality improvement. *BMC Health Serv Res.* 2020;20(1):251. DOI: 10.1186/s12913-020-5018-9
24. Kim E-Y, Cho E, June KJ. Factors influencing use of home care and nursing homes. *J Adv Nurs.* 2006;54(4):511-7. DOI: 10.1111/j.1365-2648.2006.03839.x
25. Chang M, Michelet M, Skirbekk V, Langballe EM, Hopstock LA, Sund ER, et al. Trends in the use of home care services among Norwegians 70+ and projections towards 2050: The HUNT study 1995-2017. *Scand J Caring Sci.* 2023;37(3):752-65. DOI: 10.1111/scs.13158
26. Canada G of CNRC. [Internet]. 3 de octubre de 2023. NRC emotion lexicon - NRC Publications Archive [citado 3 de octubre de 2023]. Disponible en: <https://bit.ly/46eRIQe>
27. Mohammad SM, Turney P. [Internet]. NRC Emotion Lexicon [consultado 29 de septiembre de 2023]. Disponible en: <https://bit.ly/3F1RAIH>
28. Silvestre Gómez M. Implementación de asignación jerárquica latente de Dirichlet para modelado de temas. 2018 [citado 29 de septiembre de 2023]; Recuperado: <https://bit.ly/3PZndCM>
29. Ley 14/2007, de 3 de julio, de Investigación biomédica, BOE núm. 159 [Internet]. [citado 10 de septiembre de 2023]. Recuperado: <https://www.boe.es/eli/es/l/2007/07/03/14>
30. Meza P. El posicionamiento estratégico del autor en artículos de investigación: un modelo empíricamente fundado. *Logos Rev Lingüíst Filos Lit.* 2017;27(1):152-64. DOI: 10.15443/RL2711