

## Modelado de tópicos aplicado al análisis del papel del aprendizaje automático en revisiones sistemáticas

### Modeling of topics applied to the analysis of the paper of automatic learning in systemic revisions

Andrés Mauricio Grisales-Aguirre<sup>1</sup>

Carlos Julio Figueroa-Vallejo<sup>2</sup>

**Recibido:** enero 18 de 2022

**Aceptado:** mayo 03 de 2022

#### Resumen

El objetivo de la investigación fue analizar el papel del aprendizaje automático de datos en las revisiones sistemáticas de literatura. Se aplicó la técnica de Procesamiento de Lenguaje Natural denominada modelado de tópicos, a un conjunto de títulos y resúmenes recopilados de la base de datos Scopus. Específicamente se utilizó la técnica de Asignación Latente de Dirichlet (LDA), a partir de la cual se lograron descubrir y comprender las temáticas subyacentes en la colección de documentos. Los resultados mostraron la utilidad de la técnica utilizada en la revisión exploratoria de literatura, al permitir agrupar los resultados por temáticas. Igualmente, se pudo identificar las áreas y actividades específicas donde más se ha aplicado el aprendizaje automático, en lo referente a revisiones de literatura. Se concluye que la técnica LDA es una estrategia fácil de utilizar y cuyos resultados permiten abordar una amplia colección de documentos de manera sistemática y coherente, reduciendo notablemente el tiempo de la revisión.

**Palabras clave:** modelado de tópicos, aprendizaje automático, revisiones sistemáticas, Asignación Latente de Dirichlet.

#### Abstract

The objective of the research was to analyze the role of machine data learning in systematic literature reviews. The Natural Language Processing technique called topic modeling was applied to a set of titles and abstracts collected from the Scopus database. Specifically, the Latent Dirichlet Assignment (LDA) technique was used, from which it was possible to discover and understand the underlying themes in the collection of documents. The results showed the usefulness of the technique used in the exploratory literature review, by allowing the results to be grouped by theme. Likewise, it was possible to identify the specific areas and activities where machine learning has been applied the most, in relation to literature reviews. It is concluded that the LDA technique is an easy-to-use strategy and whose results allow a wide collection of documents to be approached in a systematic and coherent manner, notably reducing the review time.

**Keywords:** topic modeling, machine learning, systematic reviews, Latent Dirichlet Allocation.

1 Matemático, Estudiante de Doctorado en Ciencias – Matemáticas, Universidad Católica Luis Amigó, Manizales, Colombia. E-mail: [andres.grisalesag@amigo.edu.co](mailto:andres.grisalesag@amigo.edu.co)

Orcid: <https://orcid.org/0000-0002-4385-4474>

2 Ingeniero de Sistemas, Especialista en Big Data e Inteligencia de Negocios, Corporación Universitaria Remington, Cauca, Colombia. E-mail: [carlos.figueroava@amigo.edu.co](mailto:carlos.figueroava@amigo.edu.co)

Orcid: <https://orcid.org/0000-0002-5793-6286>

## 1. Introducción

El propósito de las revisiones sistemáticas de literatura (RS), se puede enmarcar en los siguientes puntos: i) identificar la producción acumulada sobre una temática específica, ii) identificar los patrones y tendencias que se revelan en un área de investigación o en una temática dada, iii) ampliar la base de información a partir de nuevos hallazgos o nuevas reflexiones y iv) identificar vacíos de conocimiento a partir de los cuales se puedan plantear nuevas investigaciones (Rethlefsen *et al.*, 2021). El primer desafío que se presenta para el logro de estos objetivos es el poder seleccionar los registros relevantes de entre los miles, incluso cientos de miles, que resultan de la combinación de las palabras clave en las bases de datos (Rodríguez-Jiménez & Pérez-Jacinto, 2017).

Las mismas bases ofrecen algunas herramientas para realizar análisis básicos de los resultados de búsqueda, como los filtros y los análisis descriptivos por áreas, autores, revistas, entre otros. Otras estrategias se enfocan en el análisis de ciertas métricas asociadas tanto a los autores como a las revistas, como el índice H, muy usado en la cienciometría (Millán *et al.*, 2017). La literatura también hace referencia a estrategias como el método PRISMA, que describe una serie de pasos para seleccionar documentos relevantes en una SR (Moher *et al.*, 2014; Urrútia & Bonfill, 2010). También se dispone de tecnologías para visualizar estos resultados y profundizar en los indicadores cienciométricos, con herramientas de uso frecuente como VOSviewer (Guler *et al.*, 2021) y bibliometrix (Aria & Cuccurullo, 2017).

También vale la pena destacar estrategias como la de Tree of Science (ToS) en donde se hace uso de la red de citas y en donde, por medio de la teoría de grafos, se logra identificar aquellos trabajos con mayor cantidad de citas (raíz); los más recientes (hojas) y los estructurales (tronco), metodología que se ha usado con muy buenos resultados en campos como las ciencias económicas, la psicología y en revisiones generales de las ciencias sociales

e ingenierías (Marín *et al.*, 2017; Ramírez-Carvajal *et al.*, 2021; Ramos-Enríquez *et al.*, 2021; Valencia-Hernández *et al.*, 2020; Zuluaga *et al.*, 2016).

El principal desafío que se presenta para el uso de estas metodologías es el creciente volumen de publicaciones que surge cada año en cualquier área del conocimiento. Una de las soluciones consiste en semiautomatizar las RS haciendo uso de técnicas de aprendizaje automático, técnicas de Procesamiento de Lenguaje Natural y técnicas de aprendizaje profundo (ML, NLP y DL, por sus siglas en inglés, respectivamente) (Khamparia & Singh, 2019; Robledo *et al.*, 2021). Las ventajas del uso de estas tecnologías van desde la reducción de la carga de trabajo por parte de revisores humanos, hasta el cubrimiento de grandes volúmenes de información (Marshall & Wallace, 2019).

El ML ha sido utilizado de distintas formas de acuerdo con los objetivos de las SR (Jonlagadda *et al.*, 2015; O'Mara-Eves *et al.*, 2015). En Marshall y Wallace (2019), se presenta una estructuración de estas tareas en cuatro grandes grupos: búsqueda, selección o screening, extracción de datos y síntesis. El avance en la aplicación de estas tareas ha llevado al diseño de herramientas como Rayyan, Abastractr, ExaCT y RobotReviewer, entre muchas otras, a través de las cuales, usuarios no informáticos pueden hacer uso de estas estrategias sin tener que trabajar con los algoritmos de base que soportan estos recursos (Sutton, & Marshall, 2017).

El NLP y el DL han sido los campos de la inteligencia artificial que mayor aporte le han hecho a las tareas de extracción de datos y síntesis, siendo el DL uno de los campos con muy buenos resultados en los casos en donde la cantidad de datos es muy grande. El NLP se define como el conjunto de técnicas y estrategias que permiten que los sistemas interpreten y procesen el lenguaje humano (hablado, simbólico y escrito), complejo por naturaleza y fuente universal de gran parte del conocimiento disponible en la actualidad (Jurafsky & Martin, 2008); dentro de este campo surge el modelado de tópicos, siendo una de las técnicas que permite analizar

de manera más eficiente la estructura oculta en una colección de documentos (Gorunescu, 2011; Kowsari *et al.*, 2019).

El propósito de la presente investigación es el de aplicar el enfoque de modelado de tópicos, específicamente la técnica de Asignación Latente de Dirichlet (LDA), a un registro de búsqueda relacionado con el papel y aplicaciones del ML en las revisiones sistemáticas de literatura. A continuación, se describe la metodología empleada, junto con las generalidades del modelado LDA. Luego se describen los principales resultados junto con algunas discusiones y, finalmente, se presentan las conclusiones de este trabajo.

## 2. Metodología

### 2.1 Modelado de tópicos mediante LDA

El modelado de tópicos es una técnica dentro del campo del NLP, que permite descubrir y comprender las temáticas subyacentes en una colección de documentos. El propósito central de este método es reducir un conjunto de documentos en un marco de datos discretos, usualmente mediante una representación en vectores de dichos documentos, donde cada uno corresponde a un conteo de palabras. Una de las suposiciones básicas del modelado de tópicos es que cada documento es una mezcla aleatoria de temáticas y estas, una mezcla de palabras (Kherwa, & Bansal, 2018).

Uno de los modelos algebraicos mediante los cuales se representa un conjunto de documentos (corpus) es el de la frecuencia de término – frecuencia inversa de documento (*tf – idf*). La premisa de este modelo es que cuanto mayor sea la frecuencia de un término en un documento, más relevante va a ser dicho término para la definición de la temática del respectivo documento (*tf*). En complemento, la *idf* establece que, si un término aparece con mucha frecuencia en muchos documentos, menor será el valor discriminativo de dicho término para la definición de un tópico latente (Asmussen & Møller, 2019).

La ecuación 1 muestra la forma de calcular esta frecuencia.

$$tf - idf_{t,d} = \frac{n_{t,d}}{\sum_k n_{k,d}} \cdot \log \left( \frac{N}{df_t} \right) \quad (1)$$

El modelo de asignación latente de Dirichlet (LDA), se basa en el supuesto de que las palabras en un documento son intercambiables y cada documento se puede representar como una secuencia de palabras individuales (Blei *et al.*, 2003). Este modelo es generativo, lo que significa que analiza, en primer lugar, cómo se producen los datos y posteriormente se analiza qué variable objetivo los ha generado; las características de los tópicos y documentos se extraen de una distribución de Dirichlet, la cual corresponde a una generalización multivariada de una distribución Beta (Jelodar *et al.*, 2019). La aplicación de este modelo tiene como propósito estimar las variables latentes, es decir, los tópicos o temáticas, como una distribución condicionada a los documentos a partir de la distribución conjunta que se presenta en la ecuación 2 (Blei *et al.*, 2003).

$$P(w, z, \theta, \phi / \alpha, \beta) = P(\theta / \alpha) P(z / \theta) P(\phi / \beta) P(w / z, \phi) \quad (2)$$

En la ecuación 2 se tiene que  $P(\theta / \alpha)$  calcula la distribución de tópicos por documento a partir del parámetro de Dirichlet, el cual es un vector  $\alpha$  de componentes positivas. El cálculo de esta estimación se da mediante la ecuación 3 (Blei *et al.*, 2003).

$$P(\theta / \alpha) = \frac{\Gamma(\alpha)}{\prod_{k=1}^K \Gamma(\alpha_k)} \prod_{k=1}^K \theta_k^{\alpha_k - 1} \quad (3)$$

El operador  $\alpha$  definido en la ecuación 3, corresponde a la suma de todos los valores presentes en las variables.

También se tiene la distribución  $P(z / \theta)$  que estima la distribución del tópico  $z$  en el corpus,

esta estimación se logra mediante la ecuación 4 (Blei *et al.*, 2003).

$$P(z/\theta) = \prod_{d=1}^D \prod_{k=1}^K \theta_{d,k}^{n_{d,k}} \quad (4)$$

En este caso se estima la probabilidad para los tópicos  $z$  en todos los documentos en función del número de palabras  $n_{d,k}$  siendo este último el valor asociado a la cantidad de veces que se ha asignado la temática  $k$  a cualquier palabra del documento  $d$ .

La ecuación 2 también relaciona la probabilidad  $p(\phi/\beta)$  la cual se asocia a la distribución de términos por cada temática en el corpus. Esta estimación se obtiene de una distribución de Dirichlet con parámetro  $\beta$  y se halla con la ecuación 5 (Blei *et al.*, 2003).

$$P(\phi/\beta) = \prod_{k=1}^K \frac{\Gamma(\beta_k \cdot)}{\prod_{v=1}^V \Gamma(\beta_{k,v})} \prod_{v=1}^V \phi_{k,v}^{\beta_{k,v}-1} \quad (5)$$

En esta ecuación el parámetro  $\phi_{k,v}$  establece la probabilidad de que el término  $v$  sea obtenido cuando el tópico fuese determinado.

Finalmente, a partir de la ecuación 2 se tiene la probabilidad condicional del corpus  $w$  dado el tópico  $z$  y la distribución de las palabras  $\phi$ . Esta probabilidad se calcula mediante la ecuación 6.

$$P(w/z, \phi) = \prod_{k=1}^K \prod_{v=1}^V \phi_{k,v}^{n_{k,v}} \quad (6)$$

La ecuación 3 se evalúa utilizando el cálculo marginal de las variables latentes, buscando establecer un modelo de probabilidad específico, el corpus  $w$  y los hiperparámetros, todo esto con el fin de realizar una estimación por el método de Máxima Verosimilitud de los parámetros del modelo y la inferencia sobre los tópicos (Blei *et al.*, 2003). Lo que se ha reportado en la literatura,

es que la suma sobre todas las posibles combinaciones de asignación de temáticas conlleva a que el cálculo de esta probabilidad sea computacionalmente muy costoso. Por tanto, las aproximaciones de estas probabilidades marginales se logran utilizando algoritmos de ML (Asuncion *et al.*, 2009; Porteous *et al.*, 2008; Wang *et al.*, 2011). La evaluación de los modelos se hace utilizando métricas como la perplejidad y la coherencia, siendo la última la que más se reporta en la literatura, a partir de la cual se afirma que palabras con significados semejantes coexisten en contextos semejantes (Röder *et al.*, 2015). Esta métrica está basada en Newman *et al.* (2010) quien propuso el método de información mutua (PMI) a partir de la ecuación 7.

$$C_{UCI}(w_i, w_j) = \log \frac{p(w_i, w_j) + \epsilon}{p(w_i)p(w_j)} \quad (7)$$

donde  $p(w)$  representa la probabilidad de que  $w_i$  esté presente en un documento aleatorio y  $p(w_i, w_j)$  representa la probabilidad de que tanto  $w_i$  como  $w_j$  estén presentes en el mismo documento.

## 2.2 Fases de aplicación

La metodología empleada en este trabajo se basa en la aplicación de las siguientes fases: pre – procesado de los datos, construcción del modelo, etiquetado de tópicos y síntesis de la información. Estas tareas se implementaron mediante el lenguaje de programación Python, utilizando la librería Gensim (Prabhakaran, 2018).

La ecuación de búsqueda utilizada para recuperar los documentos iniciales fue: (TITLE-ABS-KEY (machine AND learning) AND ABS (for “systematic review”) AND NOT TITLE-ABS-KEY (“a systematic review”)). Con esta estructura se asegura que los resultados se relacionen

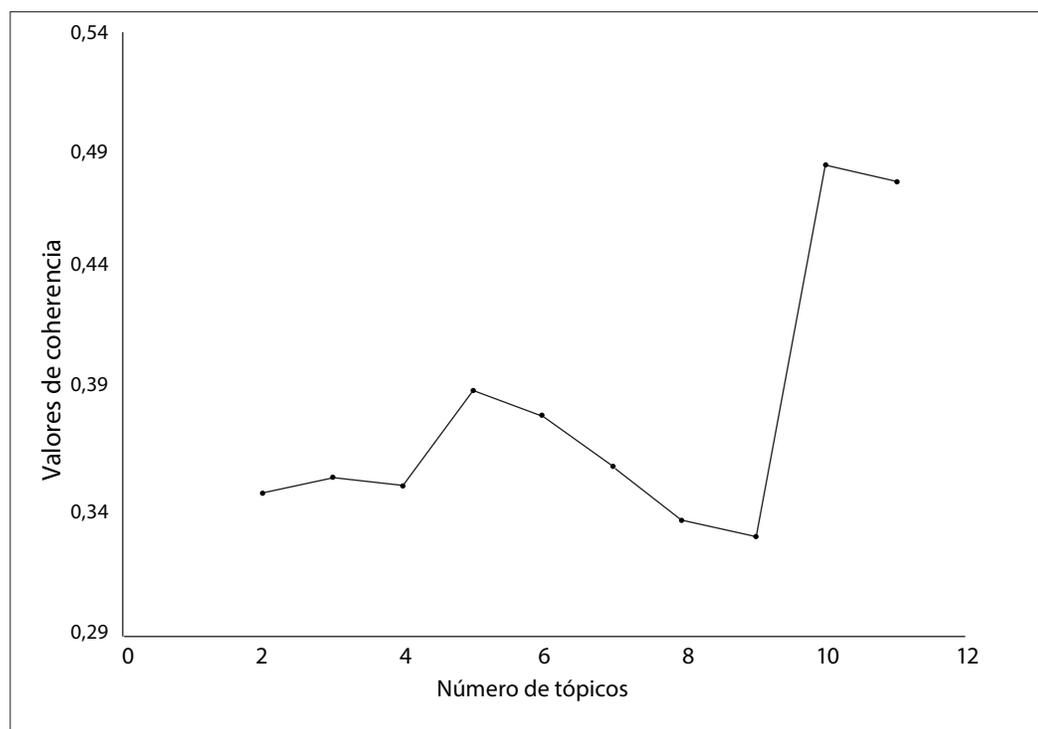
directamente con las aplicaciones del ML en las revisiones sistemáticas, en lugar de tener revisiones sistemáticas de la aplicación del ML en otras áreas. El total de registros encontrados fue de 1007 documentos, entre artículos, capítulos de libro y Reviews.

A estos registros se les aplicó una revisión preliminar para identificar los documentos directamente relacionados con la aplicación del ML en las revisiones de literatura y no en otros campos. Esta revisión se hizo utilizando la herramienta Rayyan (Ouzzani *et al.*, 2016). Esto redujo el conjunto de documentos a 202 los que se procesaron directamente con Python y la librería gensim.

### 3. Resultados y discusión

#### 3.1 Preproceso de los datos

En esta etapa se realiza la limpieza del corpus y se aplican los procesos de reducción del vocabulario (lematización, derivado o stemming, eliminación de stopwords, tokenización y definición de bigramas) (Maier *et al.*, 2018). En el presente estudio se logró reducir de 52650 términos del corpus inicial, a 35853 términos en el corpus preprocesado. El siguiente paso en el modelado de tópicos es identificar la cantidad óptima de temáticas, lo cual se evalúa a partir de los índices de coherencia. La Figura 1 muestra que con 10 tópicos se alcanza una coherencia óptima. Para esta cantidad de temáticas, el puntaje de coherencia fue de 0,4828.



**Figura 1.** Valores de coherencia para una cantidad k de tópicos.

La Tabla 1 presenta la identificación de los tópicos estimados por el modelo con sus respectivas palabras clave, el tamaño del tópico en relación al porcentaje de documentos que abarca y la eti-

queta asociada al respectivo tópico. Se presentan los tópicos que recogen la mayor proporción de documentos.

**Tabla 1.** Distribución de documentos por tópico y documento representativo.

Tópico	Palabras clave	Tamaño del tópico (%)	Etiqueta
1	machine, systematic, review, study, use, relevant, identify, learning, approach, datum	50	Clasificación automática
4	visual, screening, review, study, screen, citation, network, use, varied, separately	14	Priorización
3	review, systematic, evidence, decision_ make, update, research, automation, challenge, synthesis, base	14	Medicina basada en la evidencia
5	clinical, risk, bert, bias, trial, assessment, text, rct, patient, reviewer	10	NLP y modelado de tópicos aplicado
2	review, use, method, main, analytic, scoping, technique, objective, health, literature	6	Revisión de métodos
6	ensemble, model, article, use, high, recall, task, scientific, criterion, performance	3	Componentes de conjunto

El tópico más relevante es el número 1, al estar conformado por el 50% de los documentos del corpus. Los documentos agrupados en este conjunto están relacionados con las tareas de clasificación de documentos, particularmente resúmenes y reportes cortos, en su mayoría relacionados con el campo de las ciencias médicas (Genc *et al.*, 2020; Weißer *et al.*, 2020; Zimmerman *et al.*, 2021). En este tópico se destacan trabajos como el de Cohen *et al.* (2009), quienes determinan cómo los métodos automatizados de RS se pueden mejorar utilizando datos de entrenamiento de los modelos de otras revisiones. En el trabajo de Klymenko *et al.* (2020), se presenta

una revisión de los métodos de ML utilizados en el proceso de síntesis de texto. Otro trabajo relevante es el de Walker *et al.* (2022), donde se presentan los resultados de la aplicación de la herramienta Dextr, diseñada para extracción semiautomatizada de texto en publicaciones clínicas.

En el segundo tópico destacado, etiquetado con el número 4, asociado al proceso de priorización de documentos o screening, se encuentran trabajos que evalúan el desempeño de herramientas como: DistillerSR (Hamel *et al.*, 2020), Abstrackr (Gates, Johnson, *et al.*,

2018) o Research Screener (Chai *et al.*, 2021). También se tienen trabajos que contrastan el desempeño de estas herramientas para las mismas tareas (Gates *et al.*, 2019; Robledo *et al.*, 2021; Tsou *et al.*, 2020) y aquellos que son clave para aplicar las técnicas de ML en las revisiones sistemáticas (Hamel *et al.*, 2021; Waffenschmidt *et al.*, 2018).

En el clúster 3 se resaltan trabajos que plantean desafíos en el uso de ML en las revisiones sistemáticas de literatura. Algunos de estos trabajos plantean una estrategia que denominan *Living systematic review (LSR)*, a través de la cual se puede garantizar que se mantenga la información más reciente y actualizada. Esta estrategia se ha aplicado con mayor frecuencia en revisiones sobre la síntesis de la evidencia biomédica (EBM), donde se busca tener la mayor disponibilidad de información sobre el cuidado de pacientes (Wallace, 2018). Asimismo, se destacan trabajos como el de Elliott *et al.* (2017), quienes dan algunas indicaciones generales para el desarrollo de esta estrategia, así como en Millard *et al.* (2019) donde se pone a prueba esta estrategia en un análisis de literatura biomédica publicada en PubMed. Hablando también de la EBM, en Arno *et al.* (2021) se ponen a prueba los métodos de ML para las tareas de recuperación de registros clínicos; y en Marshall *et al.* (2020) se dan a conocer los retos para analizar la base de evidencia médica en el campo de la psicología.

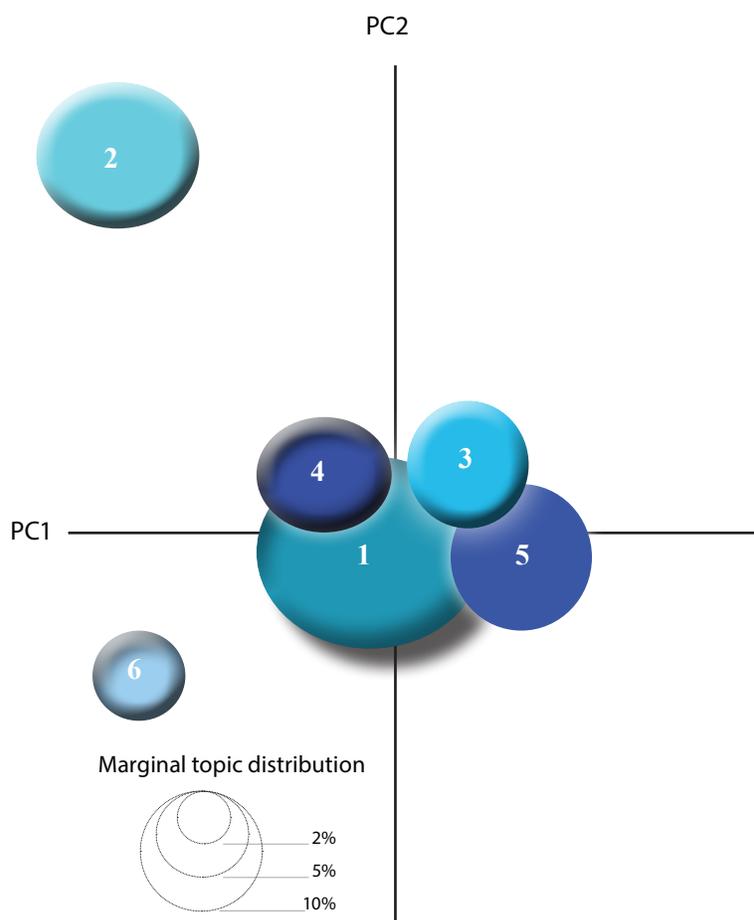
En el tópico 5 se resaltan aplicaciones de NLP y modelamiento de tópicos en revisiones de literatura médica relacionada con el dolor (Tighe *et al.*, 2020), con la evaluación de la fiabilidad de los juicios de riesgo de sesgo (Gates, Vandermeer, *et al.*, 2018) y con los ensayos aleatorios controlados (RCT) (Vinkers *et al.*, 2021). El análisis de estas aplicaciones se complementa con trabajos en donde se contrasta la eficiencia de los métodos automatizados o semiautomatizados con revisiones manuales (Soboczenski *et al.*, 2019) o con el uso de herramientas como RobotReviewer aplicadas al análisis de ensayos clínicos (Marshall *et al.*, 2016).

En el tópico 2 se destacan trabajos como el de Chishtie *et al.* (2019), donde se explora y sintetiza la literatura relacionada con la analítica visual aplicada en los servicios de investigación clínica, o el de Sangwan & Bhatnagar (2020) en donde se analizan los distintos métodos de ML que han sido aplicados en la analítica de textos. También se destaca el trabajo de Antons *et al.* (2021), donde se definen los métodos automatizados de revisión de literatura, como el de revisión computacional de literatura, y se amplía el concepto con varios ejemplos.

En el tópico 6 se tienen trabajos relacionados con métodos que combinan ML y DL, como los modelos BERT, que son arquitecturas de redes neuronales complejas y a los cuales se les pueden agregar conjuntos de características para las tareas de clasificación de documentos (Ambalavanan & Devarakonda, 2020). También se destacan trabajos en donde se ponen a prueba distintas estrategias de búsqueda con el fin de mejorar los resultados de predicción (Alamri, & Stevenson, 2015) o trabajos en donde se evalúa en términos generales el uso de herramientas de ML en las revisiones sistemáticas (Robledo *et al.*, 2021).

### 3.2 Visualización de los tópicos

Para completar el análisis del modelado de temáticas se utiliza el complemento de la librería Gensim, que permite visualizar la distribución de los tópicos en un plano bidimensional. Este resultado se basa en la aplicación de la técnica de *Multidimensional Scaling (MDS)*, la cual se aplica a la matriz de distancias entre los tópicos (ver Figura 2).



**Figura 2.** Representación bidimensional de los tópicos a través del MDS.

La visualización de los tópicos mediante este diagrama permite identificar patrones de distribución de probabilidades que se asemejan, en relación a las palabras de las temáticas identificadas. El tamaño de los círculos da una idea de la prevalencia del tópico, mientras que la distancia entre ellos da una idea de la similitud de la temática asociada.

En este caso, se observa que los tópicos identificados con los números 1, 3, 4 y 5 comparten cierta similitud entre ellos, conformando un clúster temático en relación a los sistemas de clasificación automática, su desempeño y casos de aplicación. Lo anterior se pudo concluir a partir del análisis de palabras clave y de los do-

cumentos representativos en estas temáticas. Los tópicos 2 y 6 son los que permiten definir temáticas independientes entre sí.

### Conclusiones

La primera conclusión es la posibilidad que se da de entender el protagonismo que ha venido ganando el ML en las revisiones de grandes volúmenes de registros escritos. La revisión de literatura ha sido un campo de estudio que ha madurado mucho, por un lado, debido a la necesidad latente que se tiene en cualquier proyecto de investigación de identificar el camino recorrido en el área de interés para así poder plantear

futuras investigaciones, desde una base mucho más objetiva y precisa. Por otro lado, el gran volumen de información que día a día se produce en materia de literatura científica y el desafío que se le presenta a los investigadores de contar con herramientas que les permita cubrir de manera eficiente estos grandes volúmenes de datos, ha permitido que el ML juegue un papel muy importante al desarrollar técnicas y herramientas para realizar estas tareas de revisión (Tranfield et al., 2003).

El uso de la técnica LDA permitió identificar los autores y revistas más relevantes en lo referente a la revisión de literatura científica. Esta situación es de gran relevancia para aquellas investigaciones que se desarrollen en esta misma línea o que quieran profundizar algún aspecto particular, ya que teniendo claros estos aspectos será mucho más fácil buscar directamente en estos recursos o los trabajos desarrollados por estos investigadores.

En este trabajo también se hace manifiesta la importancia y utilidad del modelado de tópicos, en particular de la técnica LDA para la revisión exploratoria de literatura científica. Se trata de una estrategia fácil de aplicar y cuyos resultados permiten abordar una amplia colección de documentos, de manera sistemática y coherente, reduciendo notablemente el tiempo utilizado para la revisión. Si bien en este trabajo no se hizo evaluación del ahorro en carga de trabajo, varias de las referencias obtenidas dan cuenta de este aspecto (Kherwa & Bansal, 2018; Qiang et al., 2020; Zhao et al., 2021).

Otro hallazgo importante fue el poder identificar en los registros de búsqueda una serie de temáticas que agrupan dichos registros, a partir de lo cual fue posible establecer cuáles son las áreas en donde se ha dado mayor demanda de este tipo de tecnologías, siendo las ciencias médicas el área en donde se encuentra mayor cantidad de evidencia empírica. Aun así, se pudo recopilar evidencia importante de aplicación de estas estrategias en otras áreas del conocimiento, donde se brindan recomendaciones valiosas para una aplicación óptima de

estas estrategias (Bertolini et al., 2021; Kang et al., 2020; Kumeno, 2020).

También se observa la utilidad de la visualización de los resultados del modelado de tópicos, lo cual permite identificar temáticas en un nivel general. Para este ejercicio, la técnica de MDS resulta apropiada y los resultados de su aplicación complementan muy bien la tarea de identificar la distribución de los documentos en temáticas y las relaciones que se puedan dar entre estas (Sami, 2020; Wei et al., 2020; Xie et al., 2018; Zhang et al., 2018).

Finalmente, se plantea como trabajo futuro el contraste de los resultados aquí reportados, limitados a Scopus, con los que se puedan obtener en otras bases de datos de alto impacto como Web of Science.

## Referencias

- Aria, M., & Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, 11(4), 959-975. <https://doi.org/10.1016/j.joi.2017.08.007>
- Alamri, A., & Stevenson, M. (2015). Automatic identification of potentially contradictory claims to support systematic reviews. *2015 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 930-937. <https://doi.org/10.1109/BIBM.2015.7359808>
- Ambalavanan, A. K., & Devarakonda, M. V. (2020). Using the contextual language model BERT for multi-criteria classification of scientific articles. *Journal of Biomedical Informatics*, 112, 103578. <https://doi.org/10.1016/j.jbi.2020.103578>
- Antons, D., Breidbach, C. F., Joshi, A. M., & Salge, T. O. (2021). Computational Literature Reviews: Method, Algorithms, and Roadmap. *Organizational Research Methods*, 1094428121991230. <https://doi.org/10.1177/1094428121991230>
- Arno, A., Elliott, J., Wallace, B., Turner, T., & Thomas, J. (2021). The views of health guideline developers on the use of automation in health evidence synthesis. *Systematic Reviews*, 10(1), 16. <https://doi.org/10.1186/s13643-020-01569-2>

- Asmussen, C. B., & Møller, C. (2019). Smart literature review: a practical topic modelling approach to exploratory literature review. *Journal of Big Data*, 6(1), 1-18. <https://doi.org/10.1186/s40537-019-0255-7>
- Asuncion, A., Welling, M., Smyth, P., & Teh, Y. W. (2009). On smoothing and inference for topic models. *Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence, UAI 2009*. (pp. 25 - 36)
- Bertolini, M., Mezzogori, D., Neroni, M., & Zammori, F. (2021). Machine Learning for industrial applications: a comprehensive literature review. *Expert Systems with Applications*, 175, 114820. <https://doi.org/10.1016/j.eswa.2021.114820>
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3, (4-5). <https://doi.org/10.1016/b978-0-12-411519-4.00006-9>
- Chai, K. E. K., Lines, R. L. J., Gucciardi, D. F., & Ng, L. (2021). Research Screener: a machine learning tool to semi-automate abstract screening for systematic reviews. *Systematic Reviews*, 10(1), 93. <https://doi.org/10.1186/s13643-021-01635-3>
- Chishtie, J. A., Babineau, J., Bielska, I. A., Cepoiu-Martin, M., Irvine, M., Koval, A., Marchand, J.-S., Turcotte, L., Jeji, T., & Jaglal, S. (2019). Visual Analytic Tools and Techniques in Population Health and Health Services Research: Protocol for a Scoping Review. *JMIR Research Protocols*, 8(10), e14019. <https://doi.org/10.2196/14019>
- Cohen, A. M., Ambert, K., & McDonagh, M. (2009). Cross-topic learning for work prioritization in systematic review creation and update. *Journal of the American Medical Informatics Association: JAMIA*, 16(5), 690-704. <https://doi.org/10.1197/jamia.M3162>
- Elliott, J. H., Synnot, A., Turner, T., Simmonds, M., Akl, E. A., McDonald, S., ... & Pearson, L. (2017). Living systematic review: 1. Introduction—the why, what, when, and how. *Journal of Clinical Epidemiology*, 91, 23-30. <https://doi.org/10.1016/j.jclinepi.2017.08.010>
- Gates, A., Guitard, S., Pillay, J., Elliott, S. A., Dyson, M. P., Newton, A. S., & Hartling, L. (2019). *Performance and Usability of Machine Learning for Screening in Systematic Reviews: A Comparative Evaluation of Three Tools*. <https://doi.org/10.23970/ahrqepc-methmachineperformance>
- Gates, A., Johnson, C., & Hartling, L. (2018). Technology-assisted title and abstract screening for systematic reviews: a retrospective evaluation of the Abstrackr machine learning tool. *Systematic Reviews*, 7(1), 45. <https://doi.org/10.1186/s13643-018-0707-8>
- Gates, A., Vandermeer, B., & Hartling, L. (2018). Technology-assisted risk of bias assessment in systematic reviews: a prospective cross-sectional evaluation of the RobotReviewer machine learning tool. *Journal of Clinical Epidemiology*, 96, 54-62. <https://doi.org/10.1016/j.jclinepi.2017.12.015>
- Genc, Y., Altuger-Genc, G., & Tatoglu, A. (2020). Systematic Review of ASEE Conference Proceedings (2007-2016) with A Machine Learning Approach. *International Journal of Engineering Education*, 36(5), 1722-1735.
- Gorunescu, F. (2011). *Data Mining: Concepts, models and techniques* (Vol. 12). Springer Science & Business Media. <https://doi.org/10.1007/978-3-642-19721-5>
- Guler, S., Capkin, S., & Sezgin, E. A. (2021). The Evolution of Publications in the Field of Scoliosis: A Detailed Investigation of Global Scientific Output Using Bibliometric Approaches. *Turkish Neurosurgery*, 31(1). <https://doi.org/10.5137/1019-5149.JTN.30216-20.2>
- Hamel, C., Hersi, M., Kelly, S. E., Tricco, A. C., Straus, S., Wells, G., Pham, B., & Hutton, B. (2021). Guidance for using artificial intelligence for title and abstract screening while conducting knowledge syntheses. *BMC Medical Research Methodology*, 21(1), 285. <https://doi.org/10.1186/s12874-021-01451-2>
- Hamel, C., Kelly, S. E., Thavorn, K., Rice, D. B., Wells, G. A., & Hutton, B. (2020). An evaluation of DistillerSR's machine learning-based prioritization tool for title/abstract screening - impact on reviewer-relevant outcomes. *BMC Medical Research Methodology*, 20(1), 256. <https://doi.org/10.1186/s12874-020-01129-1>
- Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., & Zhao, L. (2019). Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey. *Multimedia Tools and Applications*, 78(11), 15169 - 14211. <https://doi.org/10.1007/s11042-018-6894-4>
- Jonnalagadda, S. R., Goyal, P., & Huffman, M. D. (2015). Automating data extraction in systematic reviews:

- A systematic review. *Systematic Reviews*, 4(1), 1 – 16. <https://doi.org/10.1186/s13643-015-0066-7>
- Jurafsky, D., & Martin, J. H. (2008). *Speech and Language Processing: An introduction to speech recognition, computational linguistics and natural language processing*. Upper Saddle River, NJ: Prentice Hall.
- Kang, Z., Catal, C., & Tekinerdogan, B. (2020). Machine learning applications in production lines: A systematic literature review. *Computers & Industrial Engineering*, 149. <https://doi.org/10.1016/j.cie.2020.106773>
- Khamparia, A., & Singh, K. M. (2019). A systematic review on deep learning architectures and applications. *Expert Systems*, 36(3). <https://doi.org/10.1111/exsy.12400>
- Kherwa, P., & Bansal, P. (2018). Topic Modeling: A Comprehensive Review. *ICST Transactions on Scalable Information Systems*, 7(24). <https://doi.org/10.4108/eai.13-7-2018.159623>
- Klymenko, O., Braun, D., & Matthes, F. (2020). Automatic Text Summarization: A State-of-the-Art Review. En *Proceedings of the 22nd International Conference on Enterprise Information Systems*. <https://doi.org/10.5220/0009723306480655>
- Kowsari, K., Jafari-Meimandi, K., Heidarysafa, M., Mendu, S., Barnes, L., & Brown, D. (2019). Text classification algorithms: A survey. *Information*, 10(4). <https://doi.org/10.3390/info10040150>
- Kumeno, F. (2020). Software engineering challenges for machine learning applications: A literature review. *Intelligent Decision Technologies*, 13(4), 463 – 476. <https://doi.org/10.3233/idt-190160>
- Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Niekler, A., Keinert, A., Pfetsch, B., Heyer, G., Reber, U., Häussler, T., Schmid-Petri, H., & Adam, S. (2018). Applying LDA Topic Modeling in Communication Research: Toward a Valid and Reliable Methodology. *Communication Methods and Measures*, 12(2-3), 93–118. <https://doi.org/10.1080/19312458.2018.1430754>
- Marín-López, J., Robledo, S., & Duque-Méndez, N. (2017). Marketing Emprendedor: Una perspectiva cronológica utilizando Tree of Science. *Revista Civilizar De Empresa Y Economía*, 7(13), 113-123.
- Marín-Velásquez, T. D., & Arrojas-Tocuyo, D. D. J. (2021). *Revistas científicas de América Latina y el Caribe en SciELO, Scopus y Web of Science en el área de Ingeniería y Tecnología: su relación con variables socioeconómicas*. *Revista Española de Documentación Científica*, 44(3). <https://doi.org/10.3989/redc.2021.3.1812>
- Marshall, I. J., Johnson, B. T., Wang, Z., Rajasekaran, S., & Wallace, B. C. (2020). Semi-Automated evidence synthesis in health psychology: current methods and future prospects. *Health Psychology Review*, 14(1), 145–158. <https://doi.org/10.1080/17437199.2020.1716198>
- Marshall, I. J., Kuiper, J., & Wallace, B. C. (2016). RobotReviewer: evaluation of a system for automatically assessing bias in clinical trials. *Journal of the American Medical Informatics Association: JAMIA*, 23(1), 193–201. <https://doi.org/10.1093/jamia/ocv044>
- Millard, T., Synnot, A., Elliott, J., Green, S., McDonald, S., & Turner, T. (2019). Feasibility and acceptability of living systematic reviews: results from a mixed-methods evaluation. *Systematic Reviews*, 8(1), 325. <https://doi.org/10.1186/s13643-019-1248-5>
- Millán, J. D., Polanco, F., Ossa, J. C., Béria, J. S., & Cudina, J. N. (2017). La cienciometría, su método y su filosofía: Reflexiones epistémicas de sus alcances en el siglo XXI. *Revista Guillermo de Ockham*, 15(2), 17-27. <https://doi.org/10.21500/22563202.3492>
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & PRISMA, G. (2014). Ítems de referencia para publicar revisiones sistemáticas y metaanálisis: la Declaración PRISMA. *Revista Española de Nutrición Humana y Dietética*, 18(3), 172-181. <https://doi.org/10.14306/renhyd.18.3.114>
- O'Mara-Eves, A., Thomas, J., McNaught, J., Miwa, M., & Ananiadou, S. (2015). Using text mining for study identification in systematic reviews: a systematic review of current approaches. *Systematic Reviews*, 4(1), 1-22. <https://doi.org/10.1186/2046-4053-4-5>
- Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan-a web and mobile app for systematic reviews. *Systematic Reviews*, 5(1), 210. <https://doi.org/10.1186/s13643-016-0384-4>
- Porteous, I., Newman, D., Ihler, A., Asuncion, A., Smyth, P., & Welling, M. (2008). Fast collapsed gibbs sampling for latent dirichlet allocation. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, 569-577.

- Prabhakaran, S. (2018). Topic Modeling with Gensim (Python). *Machine Learning Plus*.
- Qiang, J., Qian, Z., Li, Y., Yuan, Y., & Wu, X. (2020). Short text topic modeling techniques, applications, and performance: a survey. *IEEE Transactions on Knowledge and Data Engineering*. <https://doi.org/10.1109/tkde.2020.2992485>
- Ramírez-Carvajal, D., Toro-Cardona, A., & Grisales-Aguirre, A. (2021). Competencias en networking: perspectivas desde una revisión literaria. *Revista de Ingenierías Interfaces*, 4(1), 103 -127.
- Ramos-Enríquez, V., Duque, P., & Salazar, J. A. V. (2021). Responsabilidad Social Corporativa y Emprendimiento: evolución y tendencias de investigación. *Desarrollo Gerencial*, 13(1), 1–34. <https://doi.org/10.17081/dege.13.1.4210>
- Rethlefsen, M. L., Kirtley, S., Waffenschmidt, S., Ayala, A. P., Moher, D., Page, M. J., & Koffel, J. B. (2021). PRISMA-S: an extension to the PRISMA statement for reporting literature searches in systematic reviews. *Systematic Reviews*, 10(1), 1-19. <https://doi.org/10.1186/s13643-020-01542-z>
- Robledo, S., Grisales-Aguirre, A. M., Hughes, M., & Eggers, F. (2021). “Hasta la vista, baby”—will machine learning terminate human literature reviews in entrepreneurship? *Journal of Small Business Management*, 1-30. <https://doi.org/10.1080/00472778.2021.1955125>
- Röder, M., Both, A., & Hinneburg, A. (2015). Exploring the space of topic coherence measures. In *Proceedings of the eighth ACM international conference on Web search and data mining* (pp. 399-408). <https://doi.org/10.1145/2684822.2685324>
- Rodríguez-Jiménez, A., & Pérez-Jacinto, A. O. (2017). Métodos científicos de indagación y de construcción del conocimiento. *Revista Escuela de Administración de negocios*, (82), 175-195. <https://doi.org/10.21158/01208160.n82.2017.1647>
- Sangwan, N., & Bhatnagar, V. (2020). Comprehensive Contemplation of Probabilistic Aspects in Intelligent Analytics. *International Journal of Service Science, Management, Engineering and Technology (IJSSMET)*, 11(1), 116–141. <https://doi.org/10.4018/IJSSMET.2020010108>
- Sami, I. R. (2020). Automatic Contextual Storytelling in a Natural Language Corpus. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management* (pp. 3249-3252). <https://doi.org/10.1145/3340531.3418507>
- Sutton, A., & Marshall, C. (2017). Mapping The Systematic Review Toolbox. *Value in Health*, 20(9). <https://doi.org/10.1016/j.jval.2017.08.2232>
- Soboczenski, F., Trikalinos, T. A., Kuiper, J., Bias, R. G., Wallace, B. C., & Marshall, I. J. (2019). Machine learning to help researchers evaluate biases in clinical trials: a prospective, randomized user study. *BMC Medical Informatics and Decision Making*, 19(1), 96. <https://doi.org/10.1186/s12911-019-0814-z>
- Tighe, P. J., Sannapaneni, B., Fillingim, R. B., Doyle, C., Kent, M., Shickel, B., & Rashidi, P. (2020). Forty-two Million Ways to Describe Pain: Topic Modeling of 200,000 PubMed Pain-Related Abstracts Using Natural Language Processing and Deep Learning-Based Text Generation. *Pain Medicine*, 21(11), 3133–3160. <https://doi.org/10.1093/pm/pnaa061>
- Tranfield, D., Denyer, D., & Smart, P. (2003). Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *British Journal of Management*, 14(3), 207-222. <https://doi.org/10.1111/1467-8551.00375>
- Tsou, A. Y., Treadwell, J. R., Erinoff, E., & Schoelles, K. (2020). Machine learning for screening prioritization in systematic reviews: comparative performance of Abstrackr and EPPI-Reviewer. *Systematic Reviews*, 9(1), 73. <https://doi.org/10.1186/s13643-020-01324-7>
- Urrútia, G., & Bonfill, X. (2010). Declaración PRISMA: una propuesta para mejorar la publicación de revisiones sistemáticas y metaanálisis. *Medicina clínica*, 135(11), 507-511. <https://doi.org/10.1016/j.medcli.2010.01.015>
- Valencia-Hernández, D. S., Robledo, S., Pinilla, R., Duque-Méndez, N. D., & Olivar-Tost, G. (2020). Sap algorithm for citation analysis: An improvement to tree of science. *Ingeniería e Investigación*, 40(1). <https://doi.org/10.15446/ing.investig.v40n1.77718>
- Vinkers, C. H., Lamberink, H. J., Tijdkink, J. K., Heus, P., Bouter, L., Glasziou, P., Moher, D., Damen, J. A., Hooft, L., & Otte, W. M. (2021). The methodological quality of 176,620 randomized controlled trials published between 1966 and 2018 reveals a positive trend but also an urgent need for improvement. *PLoS Biology*, 19(4), e3001162. <https://doi.org/10.1371/journal.pbio.3001162>

- Waffenschmidt, S., Hausner, E., Sieben, W., Jaschinski, T., Knelangen, M., & Overesch, I. (2018). Effective study selection using text mining or a single-screening approach: a study protocol. *Systematic Reviews*, 7(1), 166. <https://doi.org/10.1186/s13643-018-0839-x>
- Walker, V. R., Schmitt, C. P., Wolfe, M. S., Nowak, A. J., Kulesza, K., Williams, A. R., Shin, R., Cohen, J., Burch, D., Stout, M. D., Shipkowski, K. A., & Rooney, A. A. (2022). Evaluation of a semi-automated data extraction tool for public health literature-based reviews: Dextr. *Environment International*, 159, 107025. <https://doi.org/10.1016/j.envint.2021.107025>
- Wallace, B. C. (2018). Automating biomedical evidence synthesis: Recent work and directions forward. *BIRNDL@SIGIR*. <https://openreview.net/pdf?id=Hkby3SWO-B>
- Wang, C., Paisley, J., & Blei, D. (2011). Online variational inference for the hierarchical Dirichlet process. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics 752-760*. JMLR Workshop and Conference Proceedings.
- Wei, J., Han, S., & Zou, L. (2020). Vision-kg: Topic-centric visualization system for summarizing knowledge graph. In *Proceedings of the 13th International Conference on Web Search and Data Mining*, 857-860. <https://doi.org/10.1145/3336191.3371863>
- Weißer, T., Saßmannshausen, T., Ohrndorf, D., Burggräf, P., & Wagner, J. (2020). A clustering approach for topic filtering within systematic literature reviews. *MethodsX*, 7, 100831. <https://doi.org/10.1016/j.mex.2020.100831>
- Xie, T., Qin, P., & Zhu, L. (2018). Study on the Topic Mining and Dynamic Visualization in View of LDA Model. *Modern Applied Science*, 13(1), 204. <https://doi.org/10.5539/mas.v13n1p204>
- Zhang, C., Li, Z., & Zhang, J. (2018). A survey on visualization for scientific literature topics. *Journal of Visualization*, 21(2), 321-335. <https://doi.org/10.1007/s12650-017-0462-2>
- Zhao, H., Phung, D., Huynh, V., Jin, Y., Du, L., & Buntine, W. (2021). *Topic Modelling Meets Deep Neural Networks: A Survey*. <https://doi.org/10.24963/ijcai.2021/638>
- Zimmerman, J., Soler, R. E., Lavinder, J., Murphy, S., Atkins, C., Hulbert, L., Lusk, R., & Ng, B. P. (2021). Iterative guided machine learning-assisted systematic literature reviews: a diabetes case study. *Systematic Reviews*, 10(1), 97. <https://doi.org/10.1186/s13643-021-01640-6>
- Zuluaga, M., Robledo, S., Osorio-Zuluaga, G. A., Yathe, L., Gonzalez, D., & Taborda, G. (2016). Metabolomics and pesticides: systematic literature review using graph theory for analysis of references. *Nova*, 14(25), 121-138. <https://doi.org/10.22490/24629448.1735>

