

METADATOS SOCIALES: iniciativas, tecnologías, aplicaciones y softwares



Ariel Alejandro Rodríguez García
COORDINADOR



Z666.7
M48

Metadatos sociales : iniciativas, tecnologías, aplicaciones y softwares / Coordinador Ariel Alejandro Rodríguez García. - México : UNAM. Instituto de Investigaciones Bibliotecológicas y de la Información, 2024.

xv, 293 p. - (Metadatos)
ISBN: 978-607-30-8624-0

1. Metadatos. 2. Datos vinculados. 3. Indización - Aspectos sociales. 4. Tecnología de la información - Aspectos sociales. 5. Contenidos generados por los usuarios. I. Rodríguez García, Ariel Alejandro, coordinador. II. ser.

Diseño de cubierta: Mario Ocampo Chávez
Imagen: Kishore Newton - stock.adobe.com

Primera edición: Mayo de 2024
D.R. © UNIVERSIDAD NACIONAL
AUTÓNOMA DE MÉXICO
Instituto de Investigaciones Bibliotecológicas
y de la Información
Circuito Interior s/n, Torre II de Humanidades,
pisos 11, 12 y 13, Ciudad Universitaria, C. P. 04510,
Alcaldía Coyoacán, Ciudad de México

ISBN: 978-607-30-8624-0

Esta edición y sus características son propiedad de la Universidad Nacional Autónoma de México. Prohibida la reproducción total o parcial por cualquier medio sin la autorización escrita del titular de los derechos patrimoniales.

Publicación dictaminada

Impreso y hecho en México

Contenido

INTRODUCCIÓN	ix
--------------------	----

INICIATIVAS

CIUDADANÍA Y <i>DATAFICACIÓN</i> : EL ANÁLISIS SOCIOLÓGICO EN EL CONTEXTO DE LA INFORMACIÓN DIGITAL	3
Alejandro Ramos Chávez	

LOS METADATOS EN LOS PLANES DE ESTUDIO DE LOS GRADOS EN INFORMACIÓN Y DOCUMENTACIÓN: UN ENFOQUE COMPARATIVO ENTRE PORTUGAL Y ESPAÑA	17
Ana Lúcia Terra	

METADATOS EN LA FORMACIÓN PROFESIONAL EN CC II	33
Julio César Rivera Aguilera	
Luis Roberto Rivera Aguilera	
Brenda Lucero Campos Monreal	

FLUJO DE INFORMACIÓN Y USUARIOS DE REDES SOCIALES UNIVERSITARIAS: CARACTERÍSTICAS, PERFILES, NECESIDADES E IMPACTOS EN LA ORGANIZACIÓN	57
Marco Brandão	

TECNOLOGÍAS

METADATOS PARA DOCUMENTOS FÍLMICOS: INICIATIVAS Y ESTÁNDARES	77
Hilda Gabriela Lobatón Cruz	

CURACIÓN DE METADATOS PARA RECURSOS EDUCATIVOS DIGITALES	91
Ana Carolina Simionato Arakaki	

METADATOS BIBLIOGRÁFICOS Y METADATOS SOCIALES: CONEXIONES EN ENTORNOS DE DATOS VINCULADOS	113
Fabiano Ferreira de Castro	

METADATOS Y SEGURIDAD DE LA INFORMACIÓN: DESAFÍOS Y SOLUCIONES	129
Javier Moncayo García	

APLICACIONES

METADATOS SOCIALES Y PRESERVACIÓN DIGITAL: CINCO RETOS PARA LAS INSTITUCIONES DE LA MEMORIA	147
Arien González Crespo	

LOS RETOS DE LOS RECURSOS EDUCATIVOS ABIERTOS Y SU CATALOGACIÓN: CREACIÓN DE METADATOS PROFESIONALES Y SOCIALES	175
Alma Beatriz Rivera Aguilera	
Elisa Cruz Rojas	
María Guadalupe Barrera Galán	

EL SENTIDO SOCIAL DEL DATO CIENTÍFICO GENERADO POR LA BIBLIOTECA UNIVERSITARIA DESDE LA PRÁCTICA DE LA DIVULGACIÓN ACADÉMICA	199
Luisa Coral Acosta Cruz	

LA CATALOGACIÓN SOCIAL, SU PRÁCTICA PROFESIONAL Y EMPÍRICA	213
Ariel Alejandro Rodríguez García	

SOFTWARES

METODOLOGÍA PARA ESTABLECER RELACIONAMIENTO AUTOMATIZADO DE PATRONES COMUNES EN TESTIMONIOS ESCRITOS DE VÍCTIMAS DEL CONFLICTO ARMADO EN COLOMBIA	231
Fabián Orlando Baena Henao	

MODELOS Y TECNOLOGÍAS PARA LA VISUALIZACIÓN DE ONTOLOGÍAS TERMINOLÓGICAS EN EL CONTEXTO DE LA WEB SEMÁNTICA	243
Adriana Suárez Sánchez	
EL OBJETO VIRTUAL DE APRENDIZAJE (OVA) COMO PRODUCTO DE APROPIACIÓN SOCIAL DEL CONOCIMIENTO DEL BANCO DE DATOS TERMINOLÓGICOS DE LAS CIENCIAS DE LA INFORMACIÓN	265
María Teresa Múnera Torres	
APRENDIZAJE MÁQUINA EN LA BIBLIOTECOLOGÍA	277
Guadalupe Vanessa Carolina Gutiérrez Hernández Jorge Gómez Briseño	

Aprendizaje máquina en la Bibliotecología

GUADALUPE VANESSA CAROLINA GUTIÉRREZ HERNÁNDEZ

JORGE GÓMEZ BRISEÑO

Universidad Nacional Autónoma de México, México

INTRODUCCIÓN

El manejo de los datos es un tema fascinante, extremadamente útil para mirar con una perspectiva distinta el entorno, problemáticas y patrones interesantes, que sin duda resultan de interés para las humanidades, las cuales requieren comprender los fenómenos, problemas y situaciones para las disciplinas que la integran.¹ Tal como lo menciona Meneses, a diferencia de otras áreas de conocimiento, las Ciencias Sociales no visualizan los datos de manera neutral, sino como una construcción en la que ocurren diversas mediaciones, donde se pretende encontrar patrones a partir de los datos, elaborar modelos y proyectar fenómenos sociales mediante técnicas analíticas específicas, entre las que destacan el aprendizaje automatizado y la minería de datos, de las cuales se obtienen correlaciones e incluso proyecciones.²

1 María Elena Meneses Rocha, “Grandes datos, grandes desafíos para las Ciencias Sociales”, 434.

2 *ibidem*.

Metadatos sociales: iniciativas...

A continuación, se enlistan algunos ejemplos de proyectos sociales que se apoyan en el procesamiento de los datos:³

- UN Global Pulse. Iniciativa de las Naciones Unidas orientada al uso de datos, para obtener una mejor comprensión de los cambios en el bienestar humano y retroalimentación en tiempo real acerca del funcionamiento de las respuestas políticas.⁴
- UN Women. Dentro de sus planes de acción se encuentra el desarrollo de investigaciones e informes basados en datos sobre las brechas existentes, además, apoya la producción y el uso de estadísticas de género de alta calidad en la formulación de políticas. El trabajo de este grupo ha incluido el desarrollo de un conjunto mínimo de 52 indicadores de género, así como nuevas métricas para medir la violencia contra las mujeres y las niñas.⁵

Impulsado por el desarrollo de las TIC, en los últimos años, ha habido un avance en la ciencia de datos, gracias a la reducción de costos de almacenamiento de los datos, el desarrollo de algoritmos de aprendizaje máquina, así como la consolidación de los lenguajes de programación o software para el desarrollo de modelos, lo que permite dar soluciones a problemas y fortalecer la toma de decisiones.

DESARROLLO

La ciencia de datos es el método para obtener valor de diversas fuentes y grandes volúmenes de datos. Se encuentra conformado por un campo interdisciplinario que involucra matemáticas,

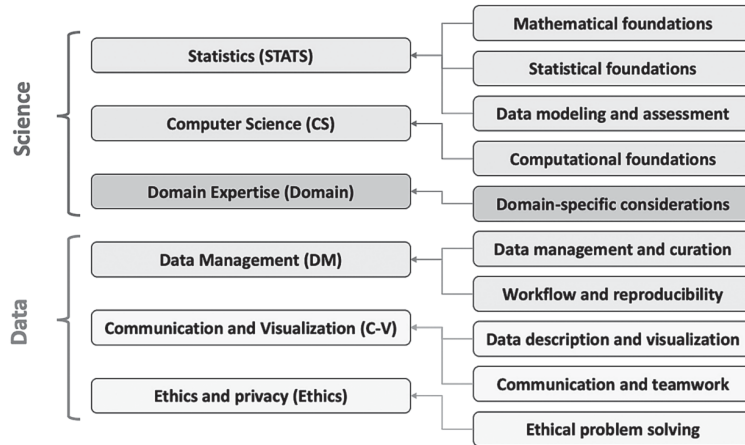
3 Massimo Lapucci y Ciro Cattuto, eds., *Data Science for Social Good*.

4 UN Global Pulse - Big Data for development and humanitarian action: <https://www.unglobalpulse.org>

5 UN Women <https://www.unwomen.org/en>

estadística, aprendizaje máquina, así como conocimiento del ámbito de estudio, entre otras habilidades.⁶ Se integra por:⁷

Figura 1. Elementos a considerar en la ciencia de datos



Fuente: Gang Shao *et al.*, *op. cit.*, 23.

- 1) Fundamentos matemáticos. Es necesaria la construcción de modelos, demostrar hipótesis y validar si estas últimas representan los problemas del mundo real.
- 2) Fundamentos estadísticos. Ciencia de recopilar, analizar, presentar e interpretar los datos.⁸
- 3) Modelado de los datos y evaluación. En este apartado se debe de identificar el modelo apropiado para el tipo de problema o pregunta de investigación.

6 Dhanurjay Patil, “A Memo to the American People from U.S. Chief Data Scientist Dr. DJ Patil: whitehouse.gov”.

7 Gang Shao *et al.*, “Exploring potential roles of academic libraries in undergraduate data science education curriculum development”, 23.

8 Facultad de Economía, “Estadística...”.

Metadatos sociales: iniciativas...

- 4) Fundamentos de computación. Habilidades para acceder y organizar datos en bases de datos, recuperarlos de los sitios web, procesamiento del lenguaje natural, administrar el almacenamiento de los datos y protegerlos desde el punto de vista informático.
- 5) Dominio del tema. Los involucrados en la aplicación de la ciencia de datos deben de tener amplio conocimiento del tema para formular las preguntas de investigación adecuadas e interpretar los resultados.
- 6) Administración y curaduría de datos. Parte fundamental de la ciencia de datos es la recuperación, preparación y acceso a los datos.
- 7) Flujo de datos y reproducibilidad. En la ciencia de datos se utilizan los flujos de datos, que muestran el camino seguido en los procesos, para resolver problemas. Documentar tales flujos permite entender el uso y análisis de los datos.
- 8) Descripción de los datos y visualización. Elemento importante para comunicar acerca de los descubrimientos del proceso de ciencia de datos. Como resultado del procesamiento, la visualización es importante para transmitir los resultados del proceso de aprendizaje máquina. La visualización es el proceso de crear imágenes que permitan entender tendencias, variaciones o descubrimiento significativo por parte de los datos.
- 9) Comunicación y trabajo de equipo. Los equipos de ciencia de datos deben formular preguntas de investigación y comunicar los resultados de manera entendible para la comunidad tanto experta como no experta. Esto implica un trabajo interdisciplinario que permita proponer soluciones creativas a diversos tipos de problemas.⁹ Otros apartados dentro de la comunicación son:

9 Engineering and Medicine National Academies of Sciences, *Data Science for Undergraduates: Opportunities...*

- Habilidad para entender las necesidades de los clientes
 - Proporcionar reportes claros
- 10) Resolución de problemas éticos. El uso potencial de los datos incrementa las posibilidades de intrusiones en la privacidad de los usuarios, por ello es importante entender el papel de la ética a través del ciclo de vida de la ciencia de datos. Este apartado se integra por:
- Principios éticos y código de conducta
 - Privacidad y confidencialidad

La Bibliotecología es una disciplina relacionada con la recopilación, administración y tratamiento de registros de información.¹⁰ La aplicación de ciencia de datos en esta rama de conocimiento tiene antecedentes importantes, entre el que destaca la propuesta del término *bibliomining*, que se define como: la minería de datos aplicada a bases de datos en bibliotecas.¹¹

APRENDIZAJE MÁQUINA

En 1950, en el artículo denominado “Computing Machinery and Intelligence”, Alan Turing se preguntaba si las computadoras podían pensar.¹² A partir de ese momento, el término de aprendizaje máquina ha ido evolucionando de acuerdo con el tiempo. En 1959 Arthur Samuel definió el aprendizaje máquina como el campo de estudio que proporciona a las computadoras la habilidad de aprender sin ser explícitamente programadas.¹³ Por su parte,

10 Tingting Zhu y Lili Zhang, “Application of data mining in the analysis of needs of university library users”.

11 Scott Nicholson, “The basis for bibliomining: Frameworks...”.

12 Alan M. Turing, “Computing Machinery and Intelligence”.

13 Arthur L. Samuel, “Some Studies in Machine Learning Using the Game of Checkers”.

Tom Mitchell¹⁴ resaltó tres elementos: la tarea, la experiencia y el desempeño, es decir, cuando una computadora ejecuta un conjunto de tareas, la experiencia debe de estar encaminada al incremento del desempeño.

Una definición más reciente indica que el aprendizaje máquina es una rama de la inteligencia artificial, y que, utilizando procesamiento computacional, es posible diseñar sistemas que puedan aprender de los datos de forma que puedan ser entrenados.

La inteligencia artificial es la teoría, métodos y tecnologías para simular inteligencia biológica en la resolución de problemas. Existe cierta interacción entre la inteligencia artificial y la ciencia de datos, pero una no es subconjunto de otra. La inteligencia artificial incluye todo lo que permite a las computadoras aprender a resolver problemas y realizar decisiones inteligentes, y la ciencia de datos se puede apoyar de dichos algoritmos.¹⁵ El aprendizaje máquina se enfoca en la detección de patrones y teoría computacional fundamentada en la inteligencia artificial.¹⁶

Dichos sistemas pueden aprender y mejorar con la experiencia y con el tiempo, y se puede refinar un modelo que puede ser usado para predecir salidas, basadas en el aprendizaje previo.¹⁷ De manera complementaria, se refiere a la construcción e implementación de un algoritmo que permita aprender de los datos y a partir de ello predecir los atributos.¹⁸

14 Tom Michael Mitchell, *Machine Learning*.

15 IBM, "What is data science?".

16 Zongben Xu *et al.*, "Data science: connotation, methods, technologies, and development".

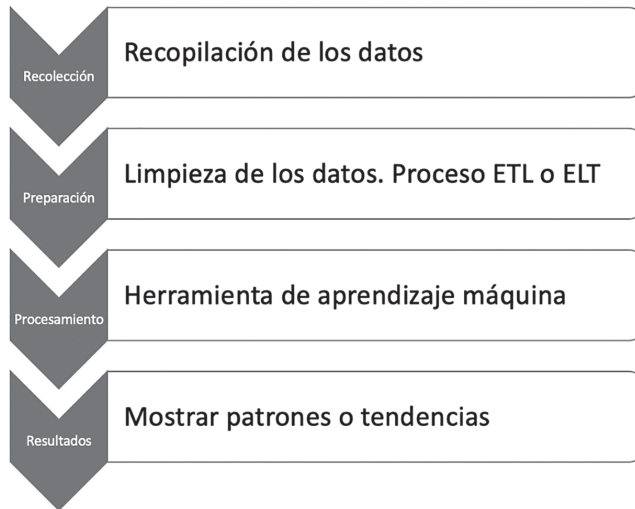
17 Jason Bell, *Machine Learning*.

18 Xu *et al.*, "Data science: connotation, methods, technologies, and development".

PROCESO DE APRENDIZAJE MÁQUINA

Si bien el término de aprendizaje máquina se refiere a la aplicación de algoritmos, es importante describir el proceso, para que dicha aplicación sea exitosa:

Figura 2. El proceso de aprendizaje máquina



Fuente: Jason Bell, *op. cit.*

- a) **Recolección.** Cuando se quiere iniciar con el uso de aprendizaje máquina, es posible preguntarse acerca de las fuentes de datos a utilizar. Actualmente existen repositorios de datos abiertos que pueden ser útiles para comenzar a experimentar.
- **Kaggle.** Es un repositorio donde es posible encontrar código y conjunto de datos aplicables a la ciencia de datos. Se integra de más de 50,000 conjuntos de datos públicos. Entre los conjuntos de datos que es posible utilizar, destacan los de la Biblioteca de San Francisco, o la MTS Library, que

contiene información de la interacción entre sus usuarios y el uso de los libros, o el del inventario de la colección de la Biblioteca Pública de Seattle. En general, Kaggle incluye 110 conjuntos de datos relacionados con bibliotecas.¹⁹

- UCI Machine Learning Repository. El repositorio de aprendizaje automático de la University of California Irvine (UCI) es una colección de bases de datos, teorías de dominio y generadores de datos que utiliza la comunidad de aprendizaje automático para el análisis de los algoritmos de aprendizaje automático.²⁰
- Datos abiertos. El portal de datos abiertos del Gobierno de la República integra el conjunto de datos de diversas instituciones gubernamentales. Los conjuntos de datos relacionados sobre bibliotecas son aproximadamente 41, entre los que es posible encontrar asistentes a bibliotecas públicas, ubicación, directorios de bibliotecas, catálogos de colecciones digitales, entre otros.²¹

Si bien los repositorios son una fuente confiable de conjunto de datos, es importante establecer las categorías de datos aplicables a las bibliotecas. La fuente de datos de la Bibliotecología se integra de una gran variedad de bases de datos, repositorios institucionales, datawarehouse, los cuales representan una gran oportunidad para la extracción de conocimiento.²² Los datos pueden ser estructurados o no estructurados y pueden componerse de las búsquedas realizadas por los usuarios en OPAC (Online Public Access Catalog), las estadísticas de uso de los recursos digitales o los movimientos de circulación de sus materiales bibliográficos, por mencionar algunos.

19 “Kaggle: Your Machine Learning and Data Science Community”.

20 “UCI Machine Learning Repository: <https://archive-beta.ics.uci.edu>.”

21 “Datos Abiertos de México”, datos.gob.mx.

22 Alexandre Ribas *et al.*, “Data science in data librarianship: Core competencies of a data librarian”.

- b) **Preparación.** Una vez identificado el conjunto de datos, el siguiente paso es prepararlo para que pueda ser usado en el aprendizaje máquina. Esta parte del proceso es fundamental, y la que en ocasiones lleva más tiempo dado que aquí se establece la calidad de los datos. Sin un conjunto de datos tratado de manera adecuada, la aplicación de herramientas de aprendizaje máquina no funcionará de manera óptima. Existen diferentes visiones para preparar los datos, entre las que destacan:
- ETL (Extract, Transform, Load - Extracción, Transformación, Carga). Es un proceso de integración de datos que combina datos de múltiples fuentes en un solo medio (normalmente un datawarehouse) u otro sistema de destino.
 - ELT (Extract, Load, Transform). La principal diferencia con ETL es el orden. ELT copia o exporta los datos de las ubicaciones de origen, pero en lugar de cargarlos en un área de preparación para la transformación, carga los datos sin procesar directamente en el datawarehouse para transformarlos según sea necesario.

Los tipos de archivo más utilizados en la extracción de datos son los siguientes:²³

- Texto en formato raw.
- Archivo separado por comas.
- JSON (Java Script Object Notation)
- XML (eXtensible Markup Language)
- Hojas de cálculo
- Bases de datos
- Imágenes

Un lenguaje de programación es una serie de instrucciones contenidas dentro de un archivo denominado código fuente y, mediante

23 Jason Bell, *Machine Learning*.

entorno tecnológico, dichas instrucciones se traducen en programas de computadora. Estos son empleados para diseñar e implementar programas encargados de definir y administrar el comportamiento de los dispositivos físicos y lógicos de un equipo de cómputo.²⁴

Existe una gran variedad de lenguajes de programación y con diferentes propósitos. Los que actualmente están enfocados a la ciencia de datos son:

- **Python.** Liderado por la Fundación de Software de Python (PSF Python Software Foundation), es un lenguaje de alto nivel, interpretado, modular, orientado a objetos. Su filosofía de lenguaje abierto permite el desarrollo de paquetes que aumentan su potencialidad. Su versión actual es la tres y es uno de los lenguajes de programación más utilizados en la ciencia de datos.
- **R.** Es un lenguaje de programación de alto nivel, usado para el cómputo estadístico, de aprendizaje máquina, y desarrollo de gráficos. Forma parte del proyecto de software libre GNU (“GNU no es UNIX”) y de la Fundación de Software Libre (Free Software Foundation FSF). Actualmente se encuentra la versión 4.2. Su nombre proviene de las iniciales de los dos primeros autores, quienes fueron Robert Gentleman y Ross Ihaka.²⁵ Usualmente R es más utilizado en la academia en áreas de la Estadística, Matemáticas y minería de datos. Entre sus características destaca la gran variedad de paquetes que permiten dirigir análisis complejos de datos y la integración óptima con otros lenguajes de programación tales como C++, Java, C., .net y Python.
- c) **Procesamiento.** Con los datos listos, es posible aplicar aprendizaje máquina. Existen diferentes tipos de algoritmos que pueden ser usados en el aprendizaje máquina, aunque

24 Francisco J. Ceballos, “Enciclopedia del lenguaje C++”.

25 Kurt Hornik y R Core Team, “R. Preguntas Frecuentes”.

su selección dependerá mucho del tipo de pregunta a contestar. Se clasifican en:

- Aprendizaje supervisado. Se refiere a trabajar con un conjunto de datos que se encuentran etiquetados. Por cada ejemplo en los datos de entrenamiento, existen campos de entrada y un campo de salida.²⁶ Un ejemplo aplicable a biblioteca sería tratar de dar respuesta a la pregunta ¿cuáles son las carreras que tienden a no regresar material bibliográfico? ¿Existe una correlación entre la carrera y el comportamiento de devolución de material bibliográfico?
 - Aprendizaje no supervisado. En este caso el algoritmo encuentra patrones ocultos de un conjunto de datos. Normalmente no existe una respuesta correcta o incorrecta, el objetivo es ejecutar el algoritmo y visualizar los posibles patrones de salida.²⁷ En este caso se podría aplicar para agrupar diferentes tipos de usuarios, por carrera, por tipo de material, y descubrir tendencias en el uso de las colecciones.
- d) **Mostrar patrones o tendencias.** Como resultado del procesamiento, la visualización es importante para transmitir los resultados del proceso de aprendizaje máquina. La visualización es el proceso de crear imágenes que permitan entender tendencias, variaciones o descubrimiento significativo por parte de los datos.²⁸

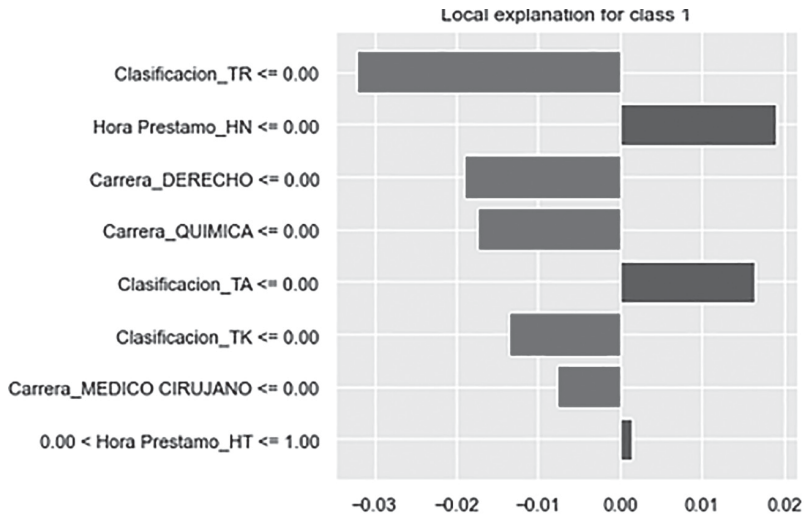
Existen diferentes formas de visualizar los datos. Dependiendo del tipo de problema, es posible seleccionar el mejor medio para representar la visualización. La siguiente imagen muestra el comportamiento de variables involucradas en el préstamo de material bibliográfico.

26 Jason Bell, *Machine Learning*.

27 Jason Bell, *idem*.

28 David Stuart, *Practical Data Science for Information Professionals*.

Figura 3. Ejemplo de representación de variables



EL ROL DEL BIBLIOTECÓLOGO EN EL APRENDIZAJE MÁQUINA

Si bien es relevante el conocimiento en disciplinas tales como Matemáticas o de computación, es igual de importante el dominio del objeto de estudio y la habilidad de contar una historia. Diversos autores coinciden en que plantearse preguntas respecto a los datos o la curiosidad para explorarlos, es indicio de un buen comienzo para obtener respuestas mediante la aplicación de un procedimiento de ciencia de datos.

El profesional de la información aplica ciencia de datos en bibliotecas mediante la implementación de esquemas para la organización de estos. De hecho, existen roles asociados al científico de datos, en los que se integra a los profesionales de la información:²⁹

29 Alma Swan y Sheridan Brown, “The skills, role and career structure of data scientists and curators: an assessment of current practice and future needs report to the JISC”.

- Creador de datos (*data creator*). Investigadores con un nivel alto del dominio del tema, productores de datos.
- Científico de datos (*data scientist*). Profesionales que trabajan en el análisis de los datos producidos por la investigación.
- Administrador de datos (*data manager*). Ingenieros en cómputo o informáticos quienes toman la responsabilidad de proveer de facilidades tecnológicas.
- Bibliotecario de datos (*data librarian*). Bibliotecólogos, especializados en la curación, preservación y almacenamiento de los datos.

El rol de bibliotecario de datos ha evolucionado y deben ser referidos como científicos de datos cuando desarrollan de forma teórica y práctica habilidades para la administración y curaduría de datos.³⁰

CONCLUSIONES

La aplicación de la ciencia de datos en la Bibliotecología es un campo interdisciplinario de amplio desarrollo. El aprendizaje máquina es la aplicación de algoritmos computacionales, pero dicha aplicación forma parte de un proceso en el que cada etapa es importante para obtener descubrimientos interesantes.

Si bien la formación del bibliotecólogo está más orientada al área de las Ciencias Sociales, existen puntos que representan un área de oportunidad:

- El aprendizaje máquina requiere la interpretación humana, con el fin de valorar los resultados proporcionados por los algoritmos. El bibliotecario de datos, como experto del dominio, tiene la capacidad de interpretar dichos patrones, así como guiar el modelo de aprendizaje máquina.

30 Alexandre Semeler, Adilson Pinto y Helen Rozados, "Data science in data librarianship: Core competencies of a data librarian".

Metadatos sociales: iniciativas...

- La aplicación de aprendizaje máquina en equipo es mejor. Normalmente los equipos de ciencia de datos, orientados a bibliotecas, consideran a especialistas en Matemáticas, Informática y expertos en el dominio de conocimiento bibliotecológico.
- Aplicar aprendizaje máquina en las bibliotecas es totalmente viable, y la mejor forma de aprender es con datos públicos o con datos privados. Además, se debe fomentar el desarrollo de más repositorios de datos, con el fin de poder realizar ciencia en dichos conjuntos y con ello replicar ejemplos relacionados al ámbito bibliotecológico.

REFERENCIAS BIBLIOGRÁFICAS

- Bell, Jason. *Machine Learning*. Indianapolis: John Wiley & Sons, Inc. 2014. <https://doi.org/10.1002/9781119183464>.
- Ceballos Sierra, Francisco Javier. *Enciclopedia del lenguaje C++*. Madrid: RA-MA. 2009, 899.
- Datos Abiertos de México - datos.gob.mx. <https://datos.gob.mx/> (Consultado el 19 de agosto de 2022).
- Dhar, Vasant. "Data science and prediction". *Communications of the ACM* 56, núm. 12 (1 de diciembre de 2013): 64–73. <https://doi.org/10.1145/2500499>.
- Facultad de Economía. "Estadística. Algunos Conceptos". <http://herzog.economia.unam.mx/profesores/blopez/eestadistica-descriptiva.pdf> (Consultado el 5 de junio de 2022).
- Hornik, Kurt y R Core Team. "R. Preguntas Frecuentes". <https://cran.r-project.org/doc/FAQ/R-FAQ.html> (Consultado el 12 de abril de 2022).
- IBM. "What is data science?" (Consultado el 25 de febrero de 2024). <https://www.ibm.com/topics/data-science>.

- “Kaggle: Your Machine Learning and Data Science Community”. <https://www.kaggle.com/> (Consultado el 19 de agosto de 2022).
- Lapucci, Massimo y Ciro Cattuto, eds. *Data Science for Social Good*. Cham: Springer International Publishing. 2021. <https://doi.org/10.1007/978-3-030-78985-5>.
- Mitchell, Tom Michael. *Machine Learning*. McGraw-Hill series in computer science. McGraw-Hill. 1997.
- National Academies of Sciences, Engineering and Medicine. *Data Science for Undergraduates: Opportunities and Options*. *Data Science for Undergraduates*. National Academies Press. 2018. <https://doi.org/10.17226/25104>.
- Nicholson, Scott. “The basis for bibliomining: Frameworks for bringing together usage-based data mining and bibliometrics through data warehousing in digital library services”. *Information Processing and Management* 42, núm. 3 (1 de mayo de 2006): 785–804. <https://doi.org/10.1016/j.ipm.2005.05.008>.
- Patil, Dhanurjay. “A Memo to the American People from U.S. Chief Data Scientist Dr. DJ Patil: whitehouse.gov”. 2015. <https://obamawhitehouse.archives.gov/blog/2015/02/19/memo-american-people-us-chief-data-scientist-dr-dj-patil>.
- Rocha, María Elena Meneses. “Grandes datos, grandes desafíos para las Ciencias Sociales”. *Revista Mexicana de Sociología* 80, núm. 2 (2018): 415–44. <https://doi.org/10.22201/iis.01882503p.2018.2.57723>.
- Samuel, Arthur L. “Some Studies in Machine Learning Using the Game of Checkers”. *IBM Journal of Research and Development* 3, núm. 3 (1959): 210–29. <https://doi.org/10.1147/rd.33.0210>.

Metadatos sociales: iniciativas...

- Semeler, Alexandre Ribas, Adilson Luiz Pinto, y Helen Beatriz Frota Rozados. "Data science in data librarianship: Core competencies of a data librarian". *Journal of Librarianship and Information Science* 51, núm. 3 (26 de septiembre de 2019): 771–80. <https://doi.org/10.1177/0961000617742465>.
- Shao, Gang, Jenny P. Quintana, Wei Zakharov, Senay Purzer, y Eunhye Kim. "Exploring potential roles of academic libraries in undergraduate data science education curriculum development". *Journal of Academic Librarianship* 47, núm. 2 (el 1 de marzo de 2021): 102320. <https://doi.org/10.1016/j.acalib.2021.102320>.
- Stuart, David. *Practical Data Science for Information Professionals*. Facet publishing. 2020.
- Swan, Alma, y Sheridan Brown. "The skills, role and career structure of data scientists and curators: an assessment of current practice and future needs report to the JISC", 2008. www.keyperspectives.co.uk.
- Turing, Alan M. "Computing Machinery and Intelligence". *Mind* 59, núm. 236 (1950): 433–60. <http://www.jstor.org/stable/2251299>.
- "UCI Machine Learning Repository". Consultado el 19 de agosto de 2022. <https://archive-beta.ics.uci.edu/>.
- "UN Global Pulse - Big Data for development and humanitarian action". Consultado el 17 de agosto de 2022. <https://www.unglobalpulse.org/>.
- "UN Women". Consultado el 17 de agosto de 2022. <https://www.unwomen.org/en>.
- Xu, Zongben, Niansheng Tang, Chen Xu, y Xueqi Cheng. "Data science: connotation, methods, technologies, and development". *Data Science and Management* 1, núm. 1 (marzo de 2021): 32–37. <https://doi.org/10.1016/j.dsm.2021.02.002>.

Zhu, Tingting, y Lili Zhang. "Application of data mining in the analysis of needs of university library users". *2011 6th International Conference on Computer Science & Education (ICCSE), Computer Science & Education (ICCSE)*, <https://doi.org/10.1109/ICCSE.2011.6028662>.

Metadatos sociales: iniciativas, tecnologías, aplicaciones y softwares. Instituto de Investigaciones Bibliotecológicas y de la Información/UNAM. La edición consta de 100 ejemplares. Coordinación editorial, Sergio J. Sepúlveda H., revisión especializada: Marcos Emilio Bustos Flores; corrección de pruebas: Carlos Ceballos Sosa, Marcos Emilio Bustos Flores; formación editorial, Mario Ocampo Chávez. Fue impreso en papel cultural de 90 g en Editorial Albatros, Av. Benito Juárez M 26 L 14, Col. El Molino Tezonco, c.p. 09960, CdMx. Se terminó de imprimir en mayo de 2024.