

Tendencias y herramientas de minería de datos para la evaluación de las publicaciones digitales**Data mining trends and tools for the evaluation of digital publications**

Córdova Vera Gonzalo Andrés, Ordoñez Ávila Ermenson Ricardo

**CIENCIA E INNOVACIÓN EN
DIVERSAS DISCIPLINAS
CIENTÍFICAS.****Julio - Diciembre, V°5-N°2;
2024**

- ✓ **Recibido:** 28/08/2024
- ✓ **Aceptado:** 06/09/2024
- ✓ **Publicado:** 31/12/2024

PAIS

- Ecuador, Portoviejo
- Ecuador, Portoviejo

INSTITUCION

- Universidad Técnica de Manabí
- Universidad Técnica de Manabí

CORREO:

- ✉ gcordova0838@utm.edu.ec
- ✉ ermenson.ordonez@utm.edu.ec

ORCID:

- <https://orcid.org/0009-0005-9868-1741>
- <https://orcid.org/0000-0003-2583-2076>

FORMATO DE CITA APA.

Córdova, G. Ordoñez, E. (2024). *Tendencias y herramientas de minería de datos para la evaluación de las publicaciones digitales*. G-ner@ndo, V°5 (N°2), 1024 – 1043.

Resumen

En la era digital actual, la accesibilidad y difusión de la información ha incrementado los desafíos relacionados con la calidad y la originalidad de las publicaciones científicas. Este problema se agrava en las instituciones de educación superior, donde la integridad académica es fundamental. Ante esto, surge la necesidad de aplicar técnicas avanzadas de minería de datos para la evaluación de publicaciones digitales, con el fin de garantizar la credibilidad y el rigor académico. El objetivo de este estudio es explorar las tendencias y herramientas de minería de datos utilizadas en la evaluación de publicaciones digitales en el ámbito académico. Para abordar este objetivo, se llevó a cabo una revisión sistemática de la literatura, guiada por el modelo PRISMA, que incluyó etapas de planificación, búsqueda y selección de estudios, y extracción de información relevante. Los resultados destacaron la implementación de herramientas como Natural Language Processing Toolkit, Scikit-learn, TensorFlow, Keras y Apache Mahout, las cuales permiten a las instituciones realizar evaluaciones más objetivas y precisas, minimizando la subjetividad en el proceso tradicional. Las conclusiones principales subrayan la importancia del uso de técnicas predictivas y de aprendizaje automático en la identificación de patrones de calidad y en la detección de plagio, contribuyendo a elevar los estándares académicos. La creciente adopción de estas tecnologías en la evaluación académica representa un avance significativo hacia la promoción de la excelencia e integridad en la investigación científica.

Palabras clave: Minería de datos, Evaluación, Publicaciones digitales, Educación superior, Aprendizaje automático.

Abstract

In the current digital era, the accessibility and dissemination of information have increased the challenges related to the quality and originality of scientific publications. This problem is aggravated in higher education institutions, where academic integrity is fundamental. Given this, the need arises to apply advanced data mining techniques for the evaluation of digital publications in order to ensure credibility and academic rigour. This study aims to explore the trends and data mining tools used in evaluating digital publications in the academic environment. A systematic literature review was conducted to address this objective, guided by the PRISMA model, which included planning stages, search and selection of studies, and extraction of relevant information. The results highlighted the implementation of tools such as Natural Language Processing Toolkit, Scikit-learn, TensorFlow, Keras and Apache Mahout, which allow institutions to perform more objective and accurate evaluations, minimizing subjectivity in the traditional process. The main conclusions underline the importance of using predictive and machine learning techniques in identifying quality patterns and plagiarism detection, contributing to raising academic standards. The growing adoption of these technologies in academic evaluation represents a significant advancement towards promoting excellence and integrity in scientific research.

Keywords: Data mining, Evaluation, Digital publishing, Higher education, Machine learning.

Introducción

Las instituciones de educación superior exigen a su cuerpo docente e investigadores cierto grado de producción científica, enfocada a diversos temas o estudios, los cuales en su mayoría son publicados de forma digital (N. M., A. D., & M. A., 2023; Ahmad et al., 2024). Estas publicaciones deben seguir un proceso de evaluación riguroso para medir su nivel de calidad, validando algunos elementos como fuentes, credenciales, novedad, autenticidad y contenido (Šakytė, 2016). Las técnicas de minería de datos son empleadas para la evaluación de publicaciones, identificando patrones, conductas, secuencias, hábitos o tendencias a través del procesamiento del texto (Khusbu & Vinit, 2021; Sulova, 2021; Xuan Lam et al., 2024).

El uso de la minería de datos ha permitido evaluar información objetiva de indicadores relacionados al contenido, aceptación, impacto o alcance de las publicaciones digitales (Mourad, Banu, Gawali, & Khaparde, 2021; Sage Research Methods Video, 2024; Sulova, 2021; Chandere, SSatish, & Lakshminarayanan, 2021). Particularmente, la minería de texto incorpora métodos que pueden facilitar la reducción de la subjetividad dentro del proceso de evaluación, elevando la calidad y prestigio académico de las mismas (Nurul, Retantyo, & SN, 2018; Dorr, Casal, & Toriano, 2021). A través del procesamiento de lenguaje natural (PLN), es posible explorar diferentes comportamientos en los datos de tipo texto (N. M. et al., 2023; Wansuree, Sorawat, & Kraisak, 2018; Rahman et al., 2024).

Para identificar los métodos y soluciones aplicadas en el proceso de validación de publicaciones digitales, es necesaria la exploración de nuevas experiencias documentadas en la literatura, que permitan describir las tendencias y la aplicación de herramientas de minería de datos más recientes (Khusbu & Vinit, 2021; Chen & Jin, 2024). Bajo esta perspectiva, el propósito de este trabajo es realizar una revisión sistemática de la literatura (SLR) que permita explorar las técnicas y herramientas de minería de datos utilizadas para la evaluación de publicaciones digitales en la educación superior.

Este trabajo está guiado por el modelo PRISMA, el cual sugiere una lista de validación de elementos principales de este tipo de documentos. De acuerdo con los lineamientos de PRISMA, el presente trabajo está organizado en los apartados de: Introducción, Métodos, Resultados, Discusión y Conclusiones.

Método y materiales

Este artículo de revisión sistemática aplicó los criterios PRISMA, organizando tres etapas: planificación, búsqueda y selección, y la extracción de la información. En investigaciones de este tipo, es esencial emplear la lista de verificación al seleccionar artículos y publicaciones, además de desarrollar una estrategia de búsqueda que se ajuste a estos criterios, todo esto se hace con el objetivo de minimizar cualquier influencia sesgada en la información recopilada.

En la fase de planificación se definió el propósito de la revisión, las preguntas de investigación y los criterios de inclusión y exclusión. Esta revisión tiene como objetivo identificar las herramientas actuales que son empleadas para la evaluación de publicaciones digitales, explicando el estado actual de hardware, software, y algoritmos de minería de datos aplicados. Para abordar este propósito, se establecieron las siguientes preguntas de investigación (RQ):

- **RQ1:** ¿Cuál es el proceso general de evaluación de publicaciones digitales?
- **RQ2:** ¿Cuáles son los criterios o parámetros seleccionados para la evaluación de publicaciones digitales?
- **RQ3:** ¿Cuáles son las soluciones tecnológicas más recientes, implementaciones de software y hardware, enfocadas en la evaluación de publicaciones digitales producidas desde la educación superior?
- **RQ4:** ¿Cuáles son los algoritmos de minería de texto con mejor rendimiento, empleados en la evaluación de publicaciones digitales?

Para la selección de los estudios se tuvieron en cuenta los títulos y resúmenes enfocados en los criterios de inclusión y exclusión (Ver Tabla 1).

Tabla 1: Criterios de Inclusión y Exclusión

Criterios de Inclusión	Criterios de Exclusión
Artículos de investigación completos	Libros, ensayos, actas de conferencia, congresos
Estudios referentes al proceso de evaluación de publicaciones digitales	Estudios duplicados
Estudios que incorporen el uso de herramientas de minería de datos, de hardware y software para la evaluación de publicaciones digitales.	Estudios que no incorporen técnicas de minado de datos, y se refieran a estadística general.
Publicaciones desde el año 2017 hasta el año 2024.	Estudios en otros idiomas diferentes al inglés y español.

Se definió un conjunto de artículos denominados grupo de control, mediante una exploración básica de trabajos que presenten información potencial para atender los requerimientos de este estudio. Estos artículos proporcionan línea base de tópicos y palabras claves que favorecen la construcción posterior de la cadena de búsqueda ideal. Para la exploración inicial de dichos artículos de referencia se utilizaron las variables de estudio para explorar de forma preliminar los términos empleados, y normalizados en la literatura (Ver Tabla 2).

Tabla 2: Artículos de grupo de control, creación de cadena de búsqueda.

#	Artículo	Palabras Claves
1	Reviewer Assignment Decision Support in an Text mining, Additives, Academic Journal based on Multicriteria Data models, Scientific publishing, Assessment and Text Mining (Latypova, 2023).	Information technology, Optimization, Nanotechnology

#	Artículo	Palabras Claves
2	Educational Data Mining (EDM) as a Model for Data mining, Electronic learning, Data Students' Evaluation in Learning Environmen (Nurul, Retantyo, & A. SN, 2018).	models, Tools, Databases, Learning management systems
3	Academic Performance Analysis Framework for Data mining, Clustering algorithms, Higher Education by Applying Data Mining Databases, Electronic learning, Techniques (Minimol & Jitendra, 2020).	Prediction algorithms, Tools
4	Tool for the evaluation of a digital scientific publication (Marbot & Roja, 2015).	Digital journals, evaluation, Evaluative tool; RECIDT journal
5	Ciberplagio académico entre el estudiantado universitario: un acercamiento al estado actual de la temática (Llovera, Aragón, & Cano, 2023).	Plagio, internet; escritura, Académica, educación superior
6	A novel technique for Thai document plagiarism detection using syntactic parse trees (Wansuree, Sorawat, & Kraisak, 2018).	Text analysis, Plagiarism detection, Natural language processing, Thai
7	Application of Text Mining Techniques on Scholarly Research Articles: Methods and Tools (Thakur & Kumar, 2021).	Knowledge discovery; Latent Dirichlet Allocation; research trends analysis; text mining; topic modelling
8	Modelos de minado de texto para la implementación de sistemas de predicción de plagio de la Universidad Técnica de Manabí (Mieles & Ricardo, 2023).	Minería de texto; predicción; plagio; software anti-plagio; publicaciones académicas.
9	Integration of LMS platform and open source algorithm for detection and prevention of Plagio académico, Plataformas de aprendizaje, Código abierto, Algoritmo	

#	Artículo	Palabras Claves
	plagiarism in Higher Education (Reduindo et al., de 2017).	detección; Ciencias de la Información

Con estas palabras claves reportadas en los artículos del grupo de control, se diseñaron cadenas de búsqueda candidatas, las mismas que fueron evaluadas en función del silencio y ruido documental de sus resultados. Finalmente, la cadena de búsqueda ideal se definió en: ("data mining" OR "digital publications") AND ("evaluation" OR "trends") AND "tools". Se realizó la búsqueda en las bases de datos IEEE Xplore, Scopus, y Science Direct. La tabla 3, presenta el formato de la cadena de búsqueda empleada en cada base de datos.

Tabla 3: Cadena de búsqueda seleccionada

Base de Datos	Cadena de Búsqueda
IEEE Xplore	("data mining" OR "digital publications") AND ("evaluation" OR "trends") AND "tools")
Scopus	TITLE-ABS-KEY (("data mining" OR "digital publications") AND ("evaluation" OR "trends") AND "tools") AND PUBYEAR > 2016 AND PUBYEAR < 2025 AND (LIMIT-TO (SUBJAREA , "COMP")) AND (LIMIT-TO (DOCTYPE , "ar")) AND (LIMIT-TO (OA , "all"))
Science direct	("data mining" OR "digital publications") AND ("evaluation" OR "trends") AND "tools"

Para la evaluación de los artículos primarios se valoraron dos aspectos principales como son el enfoque y contenido. Se evaluaron estos aspectos, en tres niveles: a) moderado, entre 0 y 40 de aportación, b) aceptable, entre 41 y 80, y c) óptimo, entre 81 y 100. El primer aspecto

permitió evaluar el enfoque de las referencias situado en trabajos de evaluación de publicaciones digitales y la minería de datos. El segundo aspecto, se centró en el contenido de calidad que tuvieron los estudios primarios, desde su estructura metodológica, hasta el aporte en limitaciones y trabajos futuros. La tabla 4 presente al detalle de los parámetros utilizados para la evaluación de artículos primarios.

Tabla 4: Criterios de evaluación de los artículos primarios

Enfoque	Contenido
E1: Mención de soluciones de hardware o software empleadas en la evaluación de las publicaciones digitales.	C1: Describe los algoritmos de minería de datos utilizados en la evaluación de publicaciones digitales.
E2: Mención de la minería de datos utilizada para la evaluación de publicaciones digitales	C2: Plantea propuestas para el desarrollo de investigaciones futuras

Extracción de la información

En la etapa de extracción de información se incluyeron criterios: enfoque del estudio, casos de uso, métodos de minería de datos y técnicas utilizadas. La Tabla 4 describe detalladamente cada uno de los criterios mencionados.

Tabla 4

Descripción de los criterios de extracción de información.

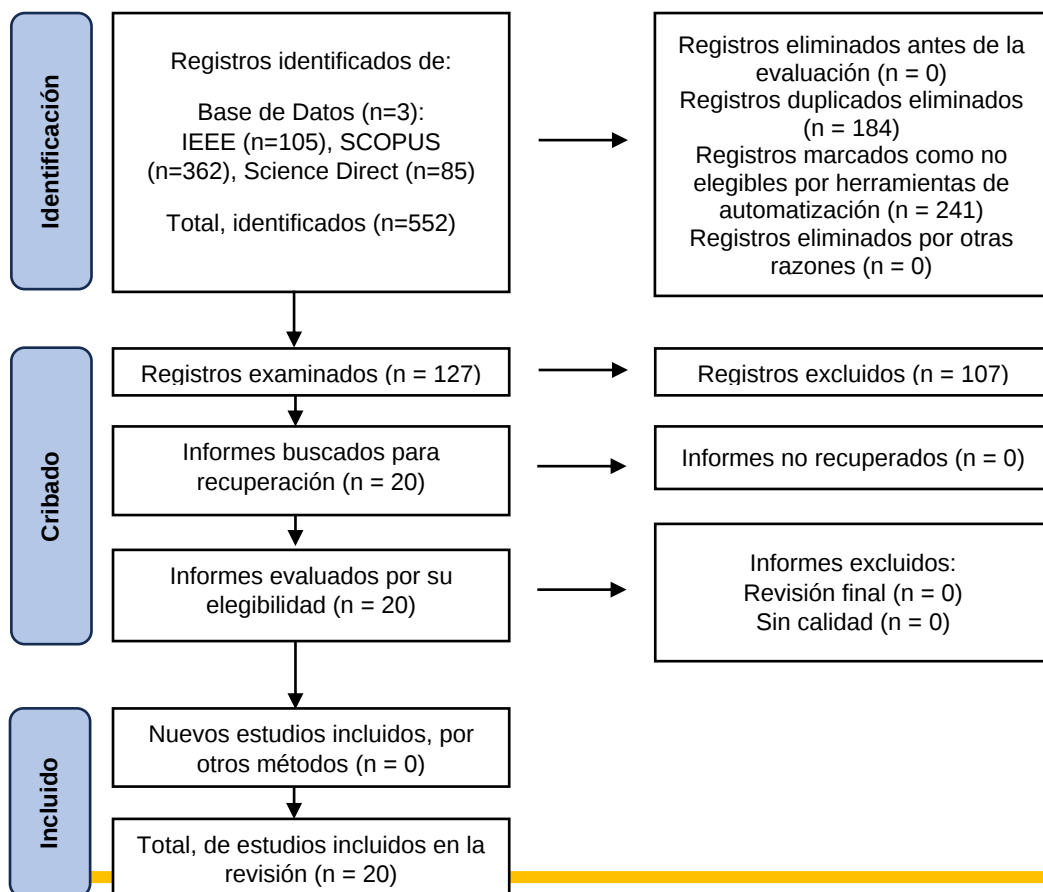
Criterios	Descripción
C1. Enfoque del estudio	Identificar el problema resuelto, el propósito de la investigación y su relación con el enfoque de este estudio.

C2. Casos de uso	Detectar casos de uso o aplicaciones de la minería de datos en la evaluación de publicaciones en la educación superior.
C3. Soluciones tecnológicas	Identificar el reporte de soluciones de software y hardware, o la tecnología utilizada en la evaluación de publicaciones digitales.
C4. Técnicas de minado de texto	Reconocer las técnicas que han sido empleadas en la evaluación de publicaciones, y su rendimiento.

Análisis de Resultados

Los principales hallazgos de este estudio mostraron que los artículos seleccionados y revisados por pares cumplieron con el 100% de los parámetros del PRISMA (Figura 1), que se realizó principalmente en 2022, cada uno con las respectivas palabras claves en relación a: evaluación de publicaciones digitales, minería de texto, minería de datos, plagio académico, etc:

Figura 1. Flujo de Búsqueda y Selección de artículos



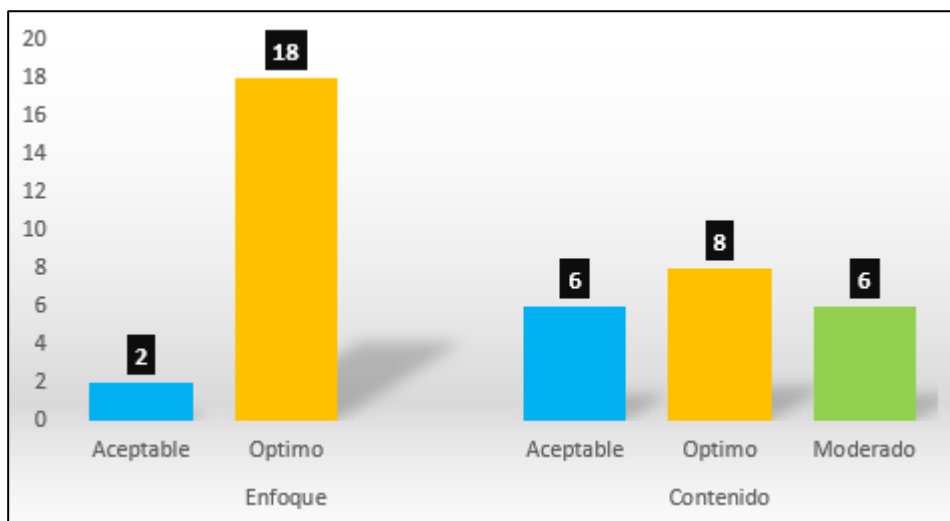
Los artículos principales seleccionados fueron evaluados según criterios específicos en cada aspecto analizado: enfoque y contenido, los cuales fueron ponderados según su relevancia en las variables de estudio (ver Tabla 7). Se destaca que hubo una contribución significativa y óptima en términos de enfoque, mientras que, en contenido, la contribución fue mayormente moderada. Al considerar los criterios de enfoque, se observó que la presentación de los indicadores de rendimiento de los algoritmos utilizados (E2) obtuvo la calificación más alta; en cuanto al contenido, el criterio más ponderado fue la metodología de minería de datos empleada (C1) (ver Figura 2).

Tabla 7: Tabla de valoración (enfoque-contenido) de artículos primarios

Artículo	Enfoque		%	Aporte	Contenido		%	Aporte
	E1	E2			C1	C2		
(Chandere et al., 2021), (N. M., A. D., & M. A., 2023)	1	1	100%	Óptimo	1	0	67%	Aceptable
(Dolader, Bel, & Muñoz, 2017)	1	0.5	85%	Óptimo	1	0	67%	Aceptable
(Dorr, Casal, & Toriano, 2021), (Wansuree, Sorawat, & Kraisak, 2018), (Mieles & Ricardo, 2023)	0.5	0.5	65%	Aceptable	0.5	0	20%	Moderado
(Khusbu & Vinit, 2021)	1	0.5	85%	Óptimo	0	1	67%	Aceptable
(Latypova, 2023), (Llovera, Aragón, & Cano, 2023), (Marbot & Roja, 2015), (Nurul, Retantyo, & A. SN,	1	1	100%	Óptimo	1	1	100%	Óptimo

Artículo	Enfoque		%	Aporte	Contenido		%	Aporte
	E1	E2			C1	C2		
2018), (Reduindo et al., 2017), (Thakur & Kumar, 2021), (Yanping & Qingyu, 2019)								
(Minimol & Jitendra, 2020)	0.5	1	85%	Óptimo	1	0	67%	Aceptable
(Mourad et al., 2021)	1	1	100%	Óptimo	0.5	0.5	35%	Moderado
(Niekler & Thomas, 2019)	1	1	100%	Óptimo	0	1	67%	Aceptable
(Novoseltseva, 2021)	0.5	1	85%	Óptimo	1	0.5	20%	Moderado
(Shah et al., 2021)	1	0.5	85%	Óptimo	0.5	0.5	35%	Moderado
(Rimpy, Dhankhar, & Solanki, 2022)	0.5	1	85%	Óptimo	1	1	100%	Óptimo

Figura 2: Nivel de aporte (enfoque-contenido)



Las técnicas principales empleadas en minería de texto para la evaluación de las publicaciones digitales incluyen clasificadores predictivos como redes neuronales, árboles de decisión, redes bayesianas, y otros basados en datos etiquetados de Machine Learning, además, del agrupamiento descriptivo, partiendo del uso del agrupamiento difuso.

En cuanto a las herramientas para la evaluación de las publicaciones digitales se puede encontrar software de predicción de plagio como son el Turnitin como la herramienta más utilizada en las instituciones de educación superior, además de otros como Chamilo y Jplag. Estas herramientas implementan características funcionales de minería de datos para la evaluación de publicaciones digitales utilizando procesos como Naive Bayes, KDD, C4.5, K-means, clasificación binaria, y soporte vectorial. Todos estos procesos se enmarcan desde el consenso, patrones de comportamiento, limpieza, extracción, validación, agrupamiento y localización de conjuntos (Tabla 7).

Tabla 7: Principales técnicas y modelos de minería de datos

Técnica de Minería de datos	Solución Funcional	Proceso	Procedimiento	Técnica de Modelo
Arboles de decisión (Mieles & Ricardo, 2023).	Dar mejor incremento a la precisión	Algoritmo XGBoost	Búsqueda de códigos fuentes plagiados	Predictivo
Redes Neuronales (Thakur & Kumar, 2021).	Mejora en el rendimiento de modelos con analices de varios impactos de datos	K-NN	Localización de palabras y conjuntos de datos copiados	Predictivo
Arboles de decisión (N. M., A. D., & M. A., 2023).	Detección de texto fraudulento	K-means, CART	Extracción de conocimiento	Descriptiva
Etiquetados de datos Machine Learning (Khusbu & Vinit, 2021).	Presencia de texto plagiado con densidad predictoras	Modelo LSTM	Clasificar y predecir	Predictiva
Arboles de análisis sintáctico (Latypova, 2023).	Mejora la precisión de la detección del texto plagiado	SRL	Etiquetado no secuencial	Predictiva

Bayesianas (Dolader, Bel, & Muñoz, 2017).	Identificación de plagio en oraciones de similitud intertextual de copia cercana	Naive Bayes	Función del consenso	Predictiva
Redes Neuronales convolucionales (Mieles & Ricardo, 2023).	Extracción de características de los objetos	Enfoque aglomerativo	Clasificar y predecir	Predictiva
Árbol de análisis comprimido (Marbot & Roja, 2015).	Clasificación de códigos fuentes	Árbol de sintaxis abstracta	Creación de árbol con diferentes nodos	Predictiva
Fuzzy Clustering (Novoseltseva, 2021).	Mejor consistencia de resultados de artículos científicos	RNN, LSTM	Jerarquía de grupos	Descriptiva
Agrupación (N. M., A. D., & M. A., 2023).	Detección de las bibliotecas digitales, como patrones comunes en documentos bajo inspección	Algoritmo ARPAD y RSA	Multivariable en estructura de datos	Descriptiva

Con esto se deduce que los indicadores del rendimiento en las técnicas de minería de datos son la precisión con una estimación de 100%, clasificación binaria con 99% y el LTSM con un 98%.

Discusión

En esta sección se discuten los hallazgos identificados en la literatura, centrados en el proceso de evaluación de las publicaciones digitales, los criterios o parámetros empleados para su evaluación, las tendencias tecnológicas utilizadas y los algoritmos de aprendizaje automático aplicados, así como sus métricas destacadas.

En atención a la pregunta RQ1: ¿Cuál es el proceso general de evaluación de publicaciones digitales? En esta investigación se pudo concluir que el proceso comienza desde la recopilación de datos, incluyendo la extracción de datos de fuentes como sitios web, bases de datos y redes sociales, entre otros, tal como lo describe N. M., A. D., y M. A. (2023). Luego se pasa a la reprocesamiento de datos, donde los datos recopilados pueden estar en diferentes formatos y pueden contener ruido o información irrelevante, según lo señalado Dorr, Casal y Toriano (2021). En esta etapa se limpian los datos, se eliminan valores atípicos, se normalizan y se transforman según sea necesario para prepararlos para el análisis, como mencionan Miele y Ricardo (2023).

Posteriormente, se realiza el análisis exploratorio de datos (EDA), llevando a cabo una exploración inicial de los datos para comprender su estructura y características, como lo plantean Dorr, Casal y Toriano (2021). Este análisis puede incluir la visualización de datos, la identificación de tendencias, la detección de correlaciones y la realización de análisis estadísticos descriptivos, de acuerdo con Miele y Ricardo (2023). Luego, se aplican técnicas de minería de datos, como la clasificación, el clustering, la regresión y la asociación, entre otras, para extraer información útil de los datos y descubrir patrones, tendencias o relaciones ocultas, tal como lo explican (N. M., A. D., & M. A., 2023) y Reduindo et al. (2017). Después de aplicar estas técnicas, es crucial evaluar la efectividad de los modelos generados, utilizando métricas de evaluación apropiadas como precisión, recall, F1-score y la matriz de confusión, de acuerdo con Chandere, Satish, y Lakshminarayanan (2021), Mourad et al. (2021), y Khusbu y Vinit (2021).

En cuanto a la pregunta RQ2: ¿Cuáles son los criterios o parámetros seleccionados para la evaluación de publicaciones digitales? Los criterios o parámetros para la evaluación pueden variar dependiendo del contexto específico y del propósito de la evaluación. Sin embargo, los criterios comunes que suelen considerarse incluyen la credibilidad de la fuente, lo cual es destacado por Latypova (2023) y Thakur y Kumar (2021). Este criterio considera la reputación y la confiabilidad de la fuente de la publicación digital, donde se evalúa la autoridad del autor o del sitio web, su historial de publicaciones, su afiliación institucional y su experiencia en el tema tratado, como lo mencionan Niekler y Thomas (2019) y Llovera, Aragón y Cano (2023).

Otro aspecto importante es la precisión de la información, lo que implica verificar la exactitud de los datos y las afirmaciones utilizadas para respaldar los argumentos presentados, tal como señalan Hernández y López (2023). Además, la actualidad es un criterio esencial, especialmente en áreas donde los cambios son frecuentes o rápidos, tal como destacan Dolader, Bel y Muñoz (2017) y Wansuree, Sorawat y Kraisak (2018). La relevancia para el tema también juega un papel fundamental, permitiendo evaluar si la publicación digital es significativa y aplicable al contexto específico, tal como lo señalan Novoseltseva (2021) y Marbot y Roja (2015).

La calidad del análisis también es crucial y está relacionada con la profundidad, coherencia y lógica del razonamiento empleado, así como con la validez de las conclusiones obtenidas, como lo señala Chandere, Satish, y Lakshminarayanan (2021). Por último, se deben tener en cuenta las referencias y fuentes citadas, evaluando la calidad y relevancia de las mismas para garantizar la fiabilidad de las fuentes secundarias y la consistencia de las citas con las normas académicas, según Llovera, Aragón y Cano (2023), Chandere, Satish, y Lakshminarayanan (2021), y Shah et al. (2021).

En cuanto a la pregunta RQ3: ¿Cuáles son las soluciones tecnológicas más recientes, implementaciones de software y hardware, enfocadas en la evaluación de publicaciones digitales producidas desde la educación superior? Actualmente, se están utilizando varias

soluciones tecnológicas en la evaluación de publicaciones digitales, como herramientas de análisis de texto basadas en inteligencia artificial y procesamiento del lenguaje natural (NLP), tal como menciona Ahmad et al. (2024). Estas herramientas pueden identificar patrones de escritura, evaluar la coherencia del argumento, detectar posibles sesgos y proporcionar retroalimentación sobre el estilo de escritura, de acuerdo con Sulova (2021). Además, los sistemas de detección de plagio juegan un papel crucial, permitiendo identificar similitudes con una amplia base de datos de documentos existentes, tal como señalan Shah et al. (2021) y Ahmad et al. (2024).

Plataformas de gestión y revisión de contenido también facilitan la creación, revisión y evaluación de publicaciones digitales, mejorando la calidad del contenido, como lo indican Minimol y Jitendra (2020). Asimismo, el análisis de redes sociales se ha vuelto relevante para evaluar el impacto y la visibilidad de las publicaciones digitales, como lo señalan Mourad et al. (2021) y Shah et al. (2021). Plataformas emergentes de aprendizaje automático están siendo implementadas para mejorar el proceso de revisión por pares, proporcionando análisis automáticos de comentarios y asignando revisores de manera eficiente, como lo destacan Dorr, Casal y Toriano (2021).

En cuanto al hardware, aunque no existen implementaciones específicas enfocadas exclusivamente en la evaluación de publicaciones digitales, el uso de sistemas de almacenamiento en la nube y servidores de alto rendimiento ha contribuido significativamente a facilitar el acceso y procesamiento de estas publicaciones en el ámbito académico, tal como lo indican Chen y Jin (2024).

Por último, respecto a la pregunta RQ4: ¿Cuáles son los algoritmos de minería de texto con mejor rendimiento empleados en la evaluación de publicaciones digitales? Entre los algoritmos más utilizados se encuentra TF-IDF, que destaca la importancia de las palabras en relación con un conjunto de documentos, según lo señalan Yanping y Qingyu (2019) y Rimpay, Dhankhar y Solanki (2022). Word2Vec también es utilizado para representar semánticamente

las palabras, permitiendo capturar relaciones sintácticas y semánticas, tal como lo mencionan Rahman et al. (2024). Otros algoritmos destacados incluyen LDA para el modelado de temas, SVM para la clasificación de textos, y BERT para la comprensión y clasificación de documentos, como lo indica Latypova (2023).

Conclusiones

Con el auge de la minería de datos en la evaluación de publicaciones digitales, se observa un creciente interés y uso de técnicas de minería de datos, especialmente la minería de texto, en la evaluación de publicaciones digitales en el ámbito de la educación superior, estas técnicas permiten analizar de manera automatizada el contenido de las publicaciones, identificar patrones, evaluar la calidad y originalidad, y mejorar los procesos de revisión y validación.

Además, con la diversidad de herramientas y técnicas utilizadas se identifican diversas herramientas y técnicas de minería de datos empleadas en la evaluación de publicaciones digitales, como Natural Language Processing Toolkit, Scikit-learn, TensorFlow, Keras y Apache Mahout. Además, se mencionan algoritmos de minería de texto con alto rendimiento, como TF-IDF, Word2Vec, LDA, SVM, RNN, BERT y FastText. Esta diversidad indica una adaptación y aplicación variada de las herramientas y algoritmos según las necesidades específicas de cada contexto de evaluación.

Y por último los desafíos y oportunidades en la evaluación de publicaciones digitales se deduce que aunque se están desarrollando soluciones tecnológicas avanzadas, como sistemas de detección de plagio y plataformas de revisión por pares basadas en aprendizaje automático, aún existen desafíos en la precisión de los recursos informáticos y en la creación de algoritmos precisos, especialmente en entornos multilingües pero sin embargo, estas tecnologías representan oportunidades significativas para mejorar la calidad y eficiencia de la evaluación de publicaciones digitales en el ámbito académico.

Referencias bibliográfica

- Ahmad, K., et al. (2024). Data-driven artificial intelligence in education: A comprehensive review. *IEEE Transactions on Learning Technologies*, 17, 12-31.
<https://doi.org/10.1109/TLT.2023.3314610>
- Chandere, V., Satish, S., & Lakshminarayanan. (2021). Online plagiarism detection tools in the digital age: A review. *Annals*.
- Chen, Y., & Jin, K. (2024). Educational performance prediction with random forest and innovative optimizers: A data mining approach. *International Journal of Advanced Computer Science and Applications*, 15(3).
<https://doi.org/10.14569/IJACSA.2024.0150308>
- Dolader, C., Bel, J., & Muñoz, J. (2017). La Blockchain: Fundamentos, aplicaciones y relación con otras tecnologías disruptivas. *Economía Industrial*.
- Dorr, R., Casal, J., & Toriano, R. (2021). Text mining in scientific publications with Argentine authors. *Medicina (Mex.)*.
- Khusbu, T., & Vinit, V. K. (2021). Application of text mining techniques on scholarly research articles: Methods and tools.
- Lam, P. X., Mai, P. Q. H., Nguyen, Q. H., Pham, T., Nguyen, T. H. H., & Nguyen, T. H. (2024). Enhancing educational evaluation through predictive student assessment modeling. *Computers & Education: Artificial Intelligence*, 6, 100244.
<https://doi.org/10.1016/j.caeai.2024.100244>
- Latypova, V. (2023). Reviewer assignment decision support in an academic journal based on multicriteria assessment and text mining. *IEEE Xplore*.
- Latypova, V. (2023). Reviewer assignment decision support in an academic journal based on multicriteria assessment and text mining. *IEEE Xplore*.
-

- Llovera, Y., Aragón, Y., & Cano, P. (2023). Ciberplagio académico entre el estudiantado universitario: un acercamiento al estado actual de la temática (2017-2020). *Revista Colombiana de Educación*.
- Marbot, E., & Roja, J. (2015). Herramienta para la evaluación de una publicación científica digital. *Redalyc*.
- Mieles, D., & Ricardo, O. (2023). Modelos de minería de texto para la implementación de sistemas de previsión de plagio en la Universidad Técnica de Manabí. *Polo del Conocimiento*.
- Minimol, A., & Jitendra, J. P. (2020). Academic performance analysis framework for higher education by applying data mining techniques. *IEEE*.
- Minimol, A., & Jitendra, J. P. (2020). Academic performance analysis framework for higher education by applying data mining techniques. *IEEE*.
- Mourad, H., Banu, S., Gawali, B., & Khaparde, V. (2021). Data mining for library service improvement by using an R tools technique: A review. *NVEO*, 22.
- N. M., A. D., & M. A. (2023). A comparative assessment of unsupervised keyword extraction tools. *IEEE Access*, 11, 144778-144798. <https://doi.org/10.1109/ACCESS.2023.3344032>
- Niekler, A., & Thomas, T. W. (2019). Text mining in evaluation. <https://d-nb.info/1176999362/34>
- Novoseltseva. (2021). Minería de datos: Casos de uso y beneficios. <https://apiumhub.com/es/tech-blog-barcelona/mineria-de-datos-casos-de-uso-beneficios/>
- Nurul, H., Retantyo, W., & A. SN. (2018). Educational data mining (EDM) as a model for students' evaluation in learning environment. *IEEE Xplore*.
- Nurul, H., Retantyo, W., & A. SN. (2018). Educational data mining (EDM) as a model for students' evaluation in learning environment. *IEEE Xplore*.
-

- Rahman, S. I., et al. (2024). Unsupervised machine learning approach for tailoring educational content to individual student weaknesses. High-Confidence Computing, 100228. <https://doi.org/10.1016/j.hcc.2024.100228>
- Reduindo, I., Rivera, L., Rivera, J., & Olvera, M. (2017). Integration of LMS platform and open-source algorithm for detection and prevention of plagiarism in higher education. Revista General de Información y Documentación.
- Rimpy, Dhankhar, A., & Solanki, K. (2022). Educational data mining tools and techniques used for prediction of student's performance: A study. In 2022 10th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO) (pp. 1-5). <https://doi.org/10.1109/ICRITO56286.2022.9965023>
- Sage Research Methods. (2024). Data science, big data analytics, and digital methods: What is data mining? Disponible en <https://methods.sagepub.com/video/what-is-data-mining>
- Šakytė, M. (2016). Lietuvos aukštojo mokslo institucijų leidžiamų mokslinių periodinių leidinių kokybės problemos. Knygotyra, 67, 191-205. <https://doi.org/10.15388/Knygotyra.67.10182>
- Shah, J., Shah, J., Baral, G., & Baral, R. (2021). Types of plagiarism and how to avoid misconduct: Pros and cons of plagiarism detection tools in research writing and publication. NJOG.
- Sulova, S. (2021). Text mining approach for identifying research trends. En Proceedings of the 22nd International Conference on Computer Systems and Technologies (pp. 93-98). ACM. <https://doi.org/10.1145/3472410.3472433>
- Thakur, K., & Kumar, V. (2021). Application of text mining techniques on scholarly research articles: Methods and tools. Routledge.
-

Wansuree, M., Sorawat, P., & Kraisak, K. (2018). A novel technique for Thai document plagiarism detection using syntactic parse trees. Engineering Applications of Science Research.

Yanping, C., & Qingyu, L. (2019). Process and application of data mining in the university library. IEEE.