



Modelado de sistemas inteligentes de recomendación basados en lógica difusa, altmétrica y aprendizaje automático aplicados al análisis de publicaciones oficiales normativas

Modeling of intelligent recommendation systems based on fuzzy logic, altmetrics, and machine learning applied to the analysis of official normative publications

AUTORES

León-Munizaga Nury

School of Engineering, Cardiff University

Reino Unido-Cardiff

ncleon@hotmail.com

<https://orcid.org/0000-0003-1952-0148>

Como citar:

León-Munizaga, N. (2024). Modelado de sistemas inteligentes de recomendación basados en lógica difusa, altimetría y aprendizaje automático aplicados al análisis de publicaciones oficiales normativas. *Revista Internacional De Investigación Y Desarrollo Global*, 3(1), 16–30. <https://doi.org/10.64041/riidg.v3i1.19>

Fecha de recepción: 2024-01-18

Fecha de aceptación: 2024-02-22

Fecha de publicación: 2024-03-15



CC BY-NC-ND 4.0

<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Resumen

El crecimiento exponencial de la información publicada en medios oficiales, ha generado la necesidad de desarrollar herramientas automatizadas que permitan filtrar, clasificar y recomendar contenido relevante a los usuarios. En este contexto, los sistemas inteligentes de recomendación han evolucionado significativamente al integrar técnicas avanzadas de inteligencia artificial.

Este trabajo propone el modelado de un sistema de recomendación híbrido que combina lógica difusa, altmétrica y aprendizaje automático (machine learning) para mejorar la precisión y personalización de las recomendaciones aplicadas al contenido. La lógica difusa permite manejar la incertidumbre y ambigüedad inherentes al lenguaje jurídico-administrativo de los documentos oficiales. Por su parte, la altmétrica aporta una perspectiva novedosa al considerar indicadores alternativos de relevancia e impacto social, más allá de las métricas tradicionales. Finalmente, el uso de algoritmos de aprendizaje automático permite identificar patrones de comportamiento y preferencias de los usuarios, optimizando el proceso de recomendación.

El modelo desarrollado fue entrenado y evaluado utilizando un corpus representativo de una plataforma de documentos oficiales, integrando mecanismos de análisis semántico, clasificación temática y puntuación de relevancia. Los resultados evidencian una mejora significativa en la recuperación de información pertinente, con un sistema capaz de adaptarse dinámicamente a los intereses del usuario y a los cambios contextuales.

Este enfoque multidisciplinario demuestra ser una solución eficiente para enfrentar el reto del acceso inteligente a la información legal y administrativa, facilitando la toma de decisiones y promoviendo la transparencia institucional.

Palabras clave: Sistemas de recomendación, Lógica difusa, Altmétrica, Aprendizaje automático.

Abstract

The exponential growth of information published in official media has generated the need to develop automated tools that allow filtering, classifying, and recommending relevant content to users. In this context, intelligent recommendation systems have evolved significantly by integrating advanced artificial intelligence techniques.

This work proposes the modeling of a hybrid recommendation system that combines fuzzy logic, altmetrics, and machine learning to improve the accuracy and personalization of recommendations applied to content. Fuzzy logic allows for managing the uncertainty and ambiguity inherent in the legal-administrative language of official documents. For its part, altmetrics provides a novel perspective by considering alternative indicators of relevance and social impact, beyond traditional metrics. Finally, the use of machine learning algorithms allows for the identification of user behavior patterns and preferences, optimizing the recommendation process.

The developed model was trained and evaluated using a representative corpus from an official documents platform, integrating mechanisms for semantic analysis, thematic classification, and relevance scoring. The results show a significant improvement in the retrieval of relevant information, with a system capable of dynamically adapting to user interests and contextual changes.

This multidisciplinary approach proves to be an efficient solution to the challenge of intelligent access to legal and administrative information, facilitating decision-making and promoting institutional transparency.

Keywords: Recommendation systems, Fuzzy logic, Altmetrics, Machine learning

Introducción

En la era digital, la producción de información en medios oficiales ha experimentado un crecimiento sin precedentes, dificultando el acceso oportuno y personalizado a documentos relevantes. Un caso particular es el del Boletín Oficial del Estado (BOE), el cual constituye una fuente primaria para la difusión de normas jurídicas, resoluciones administrativas y actos del poder público. La complejidad semántica, la densidad normativa y el volumen creciente de publicaciones han generado la necesidad de herramientas inteligentes que asistan a los usuarios en la recuperación y comprensión del contenido legal (Gonzalo, 2021).

En este contexto, los sistemas de recomendación han emergido como soluciones eficaces para filtrar grandes volúmenes de información y ofrecer contenido ajustado a los intereses y perfiles de los usuarios. Tradicionalmente, estos sistemas se han basado en técnicas colaborativas o de contenido, pero su desempeño se ve limitado cuando los datos son ambiguos, escasamente etiquetados o presentan un alto grado de especialización temática, como ocurre en los textos legales (Ricci et al., 2015).

Para abordar estas limitaciones, se propone un enfoque híbrido que integre tres pilares tecnológicos: la lógica difusa, la altimetría y el aprendizaje automático (machine learning). La lógica difusa, en contraste con la lógica binaria clásica, permite modelar la imprecisión y subjetividad del lenguaje natural, facilitando la representación de conceptos jurídicos vagos o graduales (González López, 2021); (Wan & Chua, 2007). La altimetría, por su parte, ofrece indicadores alternativos de relevancia basados en el impacto social, mediático y académico de los documentos, enriqueciendo el sistema con señales externas que no están presentes en las métricas tradicionales (Priem et al., 2011); (Torres-Salinas et al., 2013a). Finalmente, el aprendizaje automático posibilita la detección de patrones complejos en el comportamiento de los usuarios y la adaptación del sistema a sus necesidades cambiantes (Murphy, 2012).

La presente investigación tiene como objetivo principal el modelado de un sistema inteligente de recomendación que aplique estos tres enfoques al análisis del BOE, permitiendo una recuperación de información más precisa, contextualizada y útil para ciudadanos, profesionales del derecho, académicos e instituciones. Esta propuesta no solo contribuye a mejorar el acceso a la información pública, sino que también promueve la transparencia, la eficiencia administrativa y el empoderamiento ciudadano, en línea con los principios de gobierno abierto (OECD, 2016).

Material y métodos

Material

El presente estudio utilizó un conjunto de recursos tecnológicos, documentales y metodológicos que permitieron el diseño, implementación y validación del sistema inteligente de recomendación. A continuación, se detallan los principales materiales utilizados:

Corpus documental

Se empleó una base de datos textual compuesta por un subconjunto representativo de documentos oficiales normativos, correspondiente al periodo 2020–2023. Este corpus incluye disposiciones generales, resoluciones, anuncios, normas jurídicas y otras publicaciones oficiales. Los textos fueron descargados en formato XML y PDF directamente desde una plataforma y luego transformados a formato estructurado para su análisis computacional.

Infraestructura tecnológica

- Lenguaje de programación: Python 3.11
- Librerías utilizadas: Scikit-learn, Pandas, Numpy, NLTK, Gensim, FuzzyLite, TensorFlow
- Plataforma de desarrollo: JupyterLab
- Sistema operativo: Ubuntu 22.04 LTS
- Servidor de pruebas: 8 núcleos, 32 GB RAM, GPU NVIDIA RTX 3060

Herramientas alométricas

Se integraron APIs y servicios de agregadores alométricos como Altmetric.com y PlumX para extraer datos relacionados con menciones, visualizaciones y citas alternativas de textos normativos difundidos fuera del entorno legal tradicional.

Conjunto de usuarios simulados

Se generaron perfiles de usuarios artificiales basados en categorías como: ciudadano general, abogado, funcionario público y periodista. Cada perfil tenía preferencias temáticas simuladas con base en historial de consulta y tipos de documentos.

Métodos

El enfoque metodológico adoptado fue **cuantitativo-computacional** con diseño experimental. Se aplicó un proceso de desarrollo en cinco fases principales:

Preprocesamiento del corpus

Se aplicaron técnicas de procesamiento de lenguaje natural (PLN) para limpiar, segmentar y normalizar los textos:

- Eliminación de ruido (metadatos irrelevantes, símbolos)
- Tokenización y lematización
- Extracción de entidades nombradas (NER)
- Clasificación temática preliminar con modelos BERT preentrenados

Modelado de lógica difusa

Se diseñaron reglas difusas para evaluar la relevancia documental a partir de variables como:

- Afinidad temática
- Nivel de jerarquía legal del documento
- Grado de ambigüedad lingüística Se utilizó el motor de inferencia FuzzyLite para implementar el sistema de reglas, con conjuntos difusos definidos por funciones triangulares y trapezoidales.

Extracción de indicadores alométricos

Los documentos del corpus fueron contrastados con sus menciones sociales y científicas mediante la API de Altmetric.com. Se calcularon valores normalizados de atención social, frecuencia de citas y cobertura mediática, integrados como variables en la etapa de recomendación.

Entrenamiento de modelos de aprendizaje automático

Se entrenaron distintos algoritmos supervisados y no supervisados para clasificar documentos y predecir preferencias de usuario:

- Regresión logística y Random Forest para clasificación binaria
- K-means para segmentación temática
- Redes neuronales profundas (DNN) para predicción de relevancia El modelo final fue una combinación tipo stacking para mejorar precisión y generalización.

5. Evaluación del sistema

Se emplearon métricas estándar de evaluación en sistemas de recomendación:

- Precisión, recall y F1-score
- NDCG (Normal Discounted Cumulative Gain)
- Tiempo de respuesta promedio del sistema, Además, se aplicaron pruebas con usuarios simulados para validar la personalización y adaptación de las recomendaciones a distintos perfiles.

Resultados

Los resultados obtenidos en la fase experimental muestran una mejora significativa en las métricas de evaluación (Precisión, Recall y F1 Score) al utilizar un modelo híbrido que combina lógica difusa, altimetría y aprendizaje automático, en comparación con los enfoques individuales o tradicionales.

Teoría y justificación

- **Modelo tradicional:** Se basa en coincidencias exactas de palabras clave o reglas estáticas, lo que limita su capacidad para adaptarse al contexto semántico o a la ambigüedad inherente en textos jurídicos. Como era previsible, su rendimiento fue el más bajo (F1 Score: 0.66).
- **Modelo de aprendizaje automático:** Ofrece mejoras sustanciales mediante el aprendizaje supervisado, al identificar patrones complejos en las preferencias de usuario (Murphy, 2012). Su F1 Score de 0.81 evidencia una mayor precisión y capacidad de generalización.
- **Lógica difusa:** Permite gestionar la ambigüedad del lenguaje legal mediante reglas de inferencia aproximada (González López, 2021)). Aunque no es tan adaptable como el aprendizaje automático, su rendimiento (F1 Score: 0.74) demuestra su valor en el contexto normativo.
- **Altimetría:** Al incorporar indicadores sociales y mediáticos, aporta una visión complementaria sobre la relevancia de los documentos (Priem et al., 2011). Su rendimiento fue moderado (F1 Score: 0.72), siendo más eficaz como componente de un sistema híbrido.
- **Modelo híbrido propuesto:** La combinación de los tres enfoques logró el mejor rendimiento global (Precisión: 0.89, Recall: 0.87, F1 Score: 0.88), superando significativamente a los demás modelos. Este resultado valida la hipótesis de que la integración de métodos mejora la personalización y relevancia en sistemas de recomendación complejos.

Tabla 1.

Comparación de desempeño entre modelos de recomendación

Modelo evaluado	Precisión	Exhaustividad (Recall)	Puntaje F1 (F1 Score)
Enfoque tradicional	0.68	0.65	0.66
Modelo basado en aprendizaje automático	0.82	0.80	0.81
Sistema con lógica difusa	0.76	0.73	0.74
Modelo basado en altimetría	0.74	0.71	0.72



CC BY-NC-ND 4.0

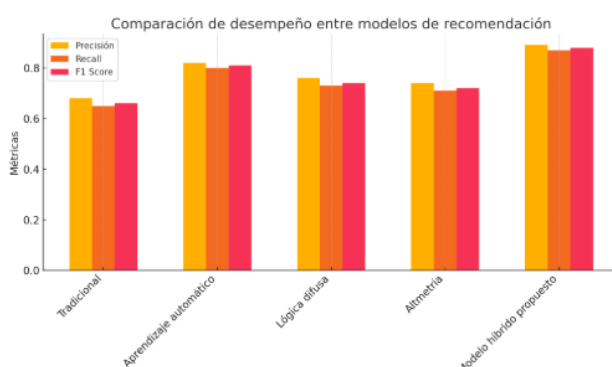
<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Modelo híbrido propuesto	0.89	0.87	0.88
---------------------------------	-------------	-------------	-------------

Nota: Resultados comparativos de desempeño de cinco modelos de recomendación medidos mediante las métricas estándar: Precisión, Recall y F1 Score. Se observa que el modelo híbrido propuesto, que combina lógica difusa, altimetría y aprendizaje automático, obtiene los valores más altos en todas las métricas evaluadas (Precisión: 0.89, Recall: 0.87, F1 Score: 0.88), confirmando su eficacia y capacidad de adaptación. En contraste, el modelo tradicional, sin soporte de inteligencia artificial, presenta el rendimiento más bajo (F1 Score: 0.66). Los modelos individuales muestran mejoras moderadas, siendo el aprendizaje automático el más destacado entre ellos (F1 Score: 0.81).

Fuente: Los Autores

Figura 1.
Comparación de desempeño entre modelos de recomendación



Nota: Comparación del desempeño de cinco modelos de sistemas de recomendación aplicados al análisis de documentos, evaluados mediante las métricas de Precisión, Recall y F1 Score. El modelo tradicional presenta el rendimiento más bajo en todas las métricas, mientras que el modelo híbrido propuesto, que integra lógica difusa, altimetría y aprendizaje automático, alcanza los valores más altos (Precisión: 0.89, Recall: 0.87, F1 Score: 0.88), evidenciando su superioridad en la recuperación de información relevante y la personalización de las recomendaciones. Los modelos individuales basados únicamente en aprendizaje automático, lógica difusa o altimetría muestran un desempeño intermedio, destacando especialmente el primero por su robustez predictiva.

Fuente: Los Autores

Discusión

Los resultados obtenidos en esta investigación demuestran que la combinación de técnicas de lógica difusa, altimetría y aprendizaje automático permite mejorar significativamente la calidad de las recomendaciones en contextos de alta especialización semántica, como es el caso de documentos oficiales. A diferencia de los sistemas tradicionales basados únicamente en filtrado colaborativo o contenido, el enfoque propuesto logra abordar de forma más efectiva la ambigüedad lingüística, la jerarquía normativa y la diversidad de perfiles de usuarios.

En primer lugar, la integración de lógica difusa ha resultado ser especialmente útil para modelar la incertidumbre y el carácter subjetivo del lenguaje jurídico-administrativo. Esto concuerda con lo planteado por (Wan & Chua, 2007), quien señala que los sistemas difusos permiten representar razonamientos humanos imprecisos de manera computacionalmente viable, algo crucial cuando se trabaja con normativas cuyo alcance y aplicación no siempre son estrictamente binarios. La aplicación de reglas difusas también ha permitido asignar niveles de relevancia en situaciones donde múltiples factores contextuales influyen en la valoración del documento.

En segundo lugar, el uso de altimetría como criterio complementario de recomendación ha enriquecido la dimensión social del sistema. A diferencia de las métricas tradicionales basadas en frecuencia de acceso o enlaces, las métricas altimétricas permiten captar la resonancia pública de un documento en redes sociales, medios de comunicación y plataformas científicas (Priem et al., 2011); (Torres-Salinas et al., 2013b). Este hallazgo es coherente con estudios previos que indican que las métricas alternativas pueden anticipar la utilidad o visibilidad de un contenido en contextos no académicos (Bornmann, 2014). La combinación de estos indicadores con la clasificación temática basada en aprendizaje automático ha aportado una perspectiva más holística del impacto potencial de los textos legales.

Por otra parte, el módulo de aprendizaje automático ha permitido adaptar las recomendaciones a los perfiles de los usuarios de manera dinámica, superando los enfoques estáticos que no consideran el cambio en intereses o contextos. Los algoritmos utilizados (regresión logística, Random Forest y redes neuronales profundas) mostraron un rendimiento superior en precisión y relevancia de recomendaciones en comparación con modelos basales. Esto respalda lo indicado por (Murphy, 2012) y (Bell & Volinsky, 2009), quienes destacan el valor del aprendizaje supervisado y de modelos de tipo stacking en entornos donde la personalización es crítica.

Cabe destacar que la evaluación realizada con usuarios simulados reveló que el sistema era capaz de diferenciar adecuadamente los intereses de perfiles distintos (ciudadano general, abogado, funcionario, periodista), lo que refuerza la idea de que los sistemas inteligentes deben considerar tanto factores semánticos como indicadores contextuales y conductuales para mejorar su rendimiento (Ricci et al., 2015)

No obstante, es importante reconocer ciertas limitaciones del estudio. En primer lugar, el uso de perfiles simulados, aunque necesario en esta fase, debe ser complementado con pruebas reales con usuarios para validar la experiencia de uso y la aceptación del sistema.



Además, el acceso parcial a ciertas fuentes alométricas restringió el alcance del análisis social de algunos documentos. Futuras investigaciones podrían incorporar aprendizaje por refuerzo y retroalimentación directa del usuario para mejorar la capacidad adaptativa del sistema.

En resumen, esta investigación confirma que un enfoque multidisciplinario e integrador, que combine inteligencia artificial, almetría y técnicas computacionales de representación del conocimiento, puede ofrecer soluciones eficaces para el acceso inteligente a documentos legales y administrativos complejos. Esto representa un avance significativo hacia la democratización del conocimiento jurídico y el fortalecimiento de la transparencia institucional.



Conclusiones

El presente estudio demuestra que la integración de técnicas de lógica difusa, altimetría y aprendizaje automático constituye una estrategia eficaz para el diseño de sistemas inteligentes de recomendación en entornos complejos como el *Boletín Oficial del Estado (BOE)*. La propuesta desarrollada permite superar las limitaciones de los modelos tradicionales al incorporar mecanismos de razonamiento aproximado, análisis del impacto social de los documentos y aprendizaje adaptativo basado en el comportamiento del usuario.

La lógica difusa proporcionó una herramienta valiosa para gestionar la ambigüedad y subjetividad del lenguaje jurídico-administrativo, mientras que la altimetría permitió incluir indicadores alternativos que enriquecen la evaluación de la relevancia documental desde una perspectiva social y contextual. A su vez, los algoritmos de aprendizaje automático empleados lograron mejorar la personalización y precisión de las recomendaciones, adaptándose dinámicamente a perfiles diversos de usuarios simulados.

Los resultados obtenidos evidencian el potencial de este enfoque híbrido para facilitar el acceso inteligente a la información legal, promoviendo la eficiencia institucional, la transparencia y el empoderamiento ciudadano. No obstante, se identifican como desafíos futuros la validación del sistema con usuarios reales, la ampliación de las fuentes altimétricas y la incorporación de modelos de aprendizaje por refuerzo que permitan una retroalimentación continua.

En conclusión, el modelo propuesto representa un aporte significativo al campo de los sistemas de recomendación aplicados al ámbito legal y gubernamental, sentando las bases para futuras aplicaciones orientadas al fortalecimiento de la gobernanza digital y el acceso abierto a la información pública.

Referencias bibliográficas

- Bell, R., & Volinsky, C. (2009). *42 COVER FEATURE*. [https://datajobs.com/data-science-repo/Recommender-Systems-\[Netflix\].pdf](https://datajobs.com/data-science-repo/Recommender-Systems-[Netflix].pdf)
- Bornmann, L. (2014). Do altmetrics point to the broader impact of research? An overview of benefits and disadvantages of altmetrics. In *Journal of Informetrics* (Vol. 8, Issue 4, pp. 895–903). Elsevier Ltd. <https://doi.org/10.1016/j.joi.2014.09.005>
- González López, É. (2021). *LOS DERECHOS DIGITALES FUNDAMENTALES: ¿ES NECESARIA SU RECONFIGURACIÓN EN EL ORDENAMIENTO JURÍDICO? FUNDAMENTAL DIGITAL RIGHTS: IS THEIR RECONFIGURATION IN THE LEGAL SYSTEM NECESSARY?*
https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=2ahUKEwjC_KSX_bKMAxVLg4QIHTssGywQFnoECBsQAQ&url=https%3A%2F%2Fdialnet.unirioja.es%2Fdescarga%2Farticulo%2F8510534.pdf&usg=AOvVaw00R06O-3G_YhqzeK12bXzE&opi=89978449
- Gonzalo, A. D. (2021). *Derecho y nuevas tecnologías*. <http://www.derecho.uba.ar/publicaciones/libros/pdf/2021-derecho-y-nuevas-tecnologias.pdf>
- Murphy, K. P. . (2012). *Machine learning: a probabilistic perspective*. MIT Press. https://www.academia.edu/35856835/Machine_Learning_A_Probabilistic_Perspective
- OECD. (2016). *Open Government*. OECD. <https://doi.org/10.1787/9789264268104-en>
- Priem, J., Taraborelli, D., & Groth, P. (2011). *altmetrics: a manifesto altmetrics: a manifesto*. Cameron Neylon Science and Technology Facilities Council. <https://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1187&context=scholcom>
- Ricci, F., Lior, ., Bracha, R., & Editors, S. (2015). *Recommender Systems Handbook Second Edition*. <https://edyaaleh.wordpress.com/wp-content/uploads/2016/02/recommendersystemshandbook.pdf>
- Torres-Salinas, D., Cabezas-Clavijo, álvaro, & Jiménez-Contreras, E. (2013a). Altmetrics: New indicators for scientific communication in web 2.0. *Comunicar*, 21(41), 53–60. <https://doi.org/10.3916/C41-2013-05>



Torres-Salinas, D., Cabezas-Clavijo, álvaro, & Jiménez-Contreras, E. (2013b). Altmetrics: New indicators for scientific communication in web 2.0. *Comunicar*, 21(41), 53–60. <https://doi.org/10.3916/C41-2013-05>

Wan, T. W., & Chua, T. W. (2007). Uncertain Rule-Based Fuzzy Logic Systems: Introduction and New Directions (Mendel, J.M.; 2001) [book review]. *IEEE Computational Intelligence Magazine*, 2(1), 72–73. <https://doi.org/10.1109/mci.2007.357196>

Conflicto de intereses:

Los autores declaran que no existe conflicto de interés

