

Algoritmos de clustering y técnicas de IA para la identificación de patrones espaciales: una revisión sistemática

Clustering algorithms and ai techniques for identifying spatial patterns: A systematic review

José Roman Castro San Agustín, Marco Alberto Mendoza Pérez, Anabelem Soberanes Martín

Resumen

El presente artículo muestra una revisión sistemática centrada en la implementación de algoritmos de clustering para el análisis de datos geospaciales y la integración de técnicas de inteligencia artificial (IA) en los sistemas de información geográfica (SIG). Se tiene como propósito destacar la importancia de la inteligencia artificial y su aplicación como herramientas para la identificación y el análisis de patrones de información en entornos geospaciales. Para alcanzar este objetivo, se emplearon métodos de investigación exploratoria, documental y descriptiva, además se realizó el análisis de diversas fuentes científicas, logrando un análisis crítico de los avances más significativos en este campo. Esta revisión sistemática se adaptó al contexto de los algoritmos de clustering y los sistemas de información geográfica. La metodología permitió una búsqueda de calidad con estudios relevantes, realizando inclusión, exclusión y evaluación crítica de las fuentes. El análisis reveló que los sistemas de información geográfica, combinados con técnicas de inteligencia artificial, pueden ofrecer una amplia gama de aplicaciones en diferentes áreas de estudio, donde la identificación de patrones espaciales se ha convertido en una herramienta clave para mejorar la comprensión de fenómenos complejos.

Palabras clave: Algoritmos de clustering; Sistemas de información geográfica (SIG); Patrones de información; Análisis de datos geospaciales; Inteligencia Artificial (IA).

José Roman Castro San Agustín

Universidad Autónoma del Estado de México | Estado de México | México | jcastros248@alumno.uaemex.mx
<https://orcid.org/0000-0001-6383-6700>

Marco Alberto Mendoza Pérez

Universidad Autónoma del Estado de México | Estado de México | México | mamendozap@uaemex.mx
<https://orcid.org/0000-0003-4911-4757>

Anabelem Soberanes Martín

Universidad Autónoma del Estado de México | Estado de México | México | asoberanesm@uaemex.mx
<https://orcid.org/0000-0002-1101-8279>

Abstract

The present article shows a systematic review focused on the implementation of clustering algorithms for geospatial data analysis and the integration of artificial intelligence techniques into geographic information systems. The aim is to highlight the importance of artificial intelligence and its application as tools for identifying and analyzing information patterns in geospatial environments. To achieve this objective, exploratory, documentary, and descriptive research methods were employed, along with an analysis of various scientific sources, enabling a critical evaluation of the most significant advancements in this field. This systematic review was adapted to the context of clustering algorithms and geographic information systems. This methodology ensured a high-quality search for relevant studies, including processes for inclusion, exclusion, and critical evaluation of sources. The analysis revealed that geographic information systems, when combined with artificial intelligence techniques, offer a wide range of applications across various fields of study, where the identification of spatial patterns has become a key tool for improving the understanding of complex phenomena.

Keywords: Clustering Algorithms, Geographic Information Systems (GIS), Information Patterns, Geospatial Data Analysis, Artificial Intelligence.

Introducción

El análisis de datos geoespaciales ha tomado relevancia en diversos campos de aplicación como la planificación urbana, la gestión de los recursos naturales, la gestión medioambiental, la planificación de la respuesta a emergencias, la agricultura, la planificación del transporte, la planificación de la ubicación de empresas y la geología o física (Pedada, 2013). Los algoritmos de clustering se han convertido en una herramienta fundamental para identificar patrones en entornos espaciales, permitiendo procesar grandes volúmenes de datos y extraer información relevante para la toma de decisiones.

El clustering es un proceso que emula la forma en que los seres humanos interpretan datos visuales, es importante en el aprendizaje automático y se aplica en diversas disciplinas que manejan datos. Su importancia se evidencia en áreas como la ciberseguridad, el procesamiento de documentos, la detección de fraudes y la biología molecular, destacando su valor para explorar y comprender patrones relevantes (Mohammad, 2023).

En el análisis de datos geoespaciales, la medición de la similitud entre puntos es un elemento fundamental para el funcionamiento de los algoritmos de clustering. Una de las medidas más utilizadas es la distancia euclidiana, la cual expresa la separación directa entre dos puntos en un espacio cartesiano. Este tipo de distancia, basada en el teorema de Pitágoras, permite cuantificar qué tan próximos o lejanos se encuentran los elementos dentro de un espacio bidimensional o en espacios de mayor dimensión. Su uso es común en numerosos algoritmos de agrupamiento debido a su simplicidad, interpretabilidad y aplicabilidad en diferentes contextos analíticos (Amat, 2017).

Las técnicas de aprendizaje profundo, especialmente los algoritmos de clustering, han emergido como herramientas importantes para el análisis de grandes volúmenes de datos espaciales y así tener la capacidad de revelar patrones ocultos en la información.

En la actualidad, el desarrollo de tecnologías basadas en algoritmos de clustering han tomado relevancia en distintos enfoques, como la segmentación de clientes, donde permiten agrupar consumidores según su comportamiento o características demográficas, facilitando la creación de campañas de marketing personalizadas y estrategias de precios específicas para cada grupo.

En el análisis de mercado, estos algoritmos ayudan a identificar nichos o tendencias emergentes, impulsando la innovación y adaptación de productos y servicios. En la detección de fraudes, se utilizan para analizar transacciones financieras y detectar patrones sospechosos, ayudando a prevenir actividades fraudulentas.

Algoritmos de clustering

Los algoritmos de clustering han demostrado ser útiles en la optimización de la cadena de suministro, donde agrupan ubicaciones o proveedores similares, mejorando la eficiencia logística. Finalmente, en la investigación y desarrollo de productos, el clustering permite identificar preferencias y tendencias de consumo, guiando el diseño de productos hacia soluciones más ajustadas a las necesidades del mercado (Arregui, 2024).

Este enfoque tecnológico propone una evolución significativa en los sistemas actuales de información geográfica, integrando soluciones avanzadas que potencian la gestión de datos y el análisis espacial. El objetivo principal del clustering es medir la similitud entre observaciones para crear grupos, dependiendo del tipo de información, se emplean diferentes tipos de algoritmos.

Clustering jerárquico (Hierarchical Clustering)

El clustering jerárquico se define como un método para agrupar datos en función de su distancia. El propósito es lograr que los datos dentro de un cluster sean lo más parecidos entre sí. Este enfoque puede seguir dos variantes: aglomerativo, que comienza considerando cada punto de datos como un cluster individual y luego agruparlos en conjuntos más grandes; y divisivo, que parte de un solo cluster y lo divide progresivamente en otros más pequeños (Castillo, 2023).

Los métodos jerárquicos son ampliamente utilizados debido a su capacidad para generar una jerarquía de particiones a partir de un conjunto de datos. El proceso aglomerativo fusiona pares de objetos o clusters en conjuntos mayores hasta que solo queda un único cluster. Por el contrario, el método divisivo subdivide los clusters existentes hasta que cada objeto pertenece a su propio cluster (Vichi et al., 2022).

Mean Shift

El algoritmo Mean Shift es un método iterativo cuya finalidad es localizar áreas de alta densidad de puntos de datos. Este enfoque utiliza el concepto de Kernel de Estimación de Densidad

(KDE), que asume que los puntos provienen de una distribución de probabilidad desconocida y estima esa distribución subyacente.

El algoritmo consiste en inicializar una ventana circular deslizante centrada en un punto aleatorio y, en cada iteración, desplazar esa ventana hacia regiones de mayor densidad. Se sigue moviendo hasta que se alcanza la convergencia, es decir, hasta que no se detecta un aumento significativo de puntos dentro de la ventana. Finalmente, los puntos se agrupan en función de la ventana deslizante en la que residan (Sancho, 2023).

K-Means

El algoritmo K-Means se basa en dividir un conjunto de datos en un número determinado de grupos, definidos como “K”. Se debe minimizar la suma de las distancias entre los puntos de datos y el centro de cada grupo, o centroide. El algoritmo asigna cada punto al cluster con el centroide más cercano, recalculando continuamente los centroides hasta que se estabilizan y no varían más.

El proceso comienza determinando el número de clusters deseados, seleccionando k puntos al azar como centroides iniciales, y asignando los datos a su cluster correspondiente. Posteriormente, se recalculan los centroides y se repite el ciclo hasta la convergencia (Ramírez, 2023).

DBSCAN

El algoritmo DBSCAN (Density-Based Spatial Clustering of Applications with Noise) es un enfoque que identifica clusters basándose en la densidad de los datos. A diferencia de K-Means, DBSCAN no requiere definir el número de clusters previamente, sino que emplea dos parámetros clave: ϵ (eps), que define la distancia máxima para considerar que dos puntos pertenecen al mismo cluster, y el número mínimo de muestras (min_samples), que establece la cantidad mínima de puntos cercanos para formar un cluster. Este algoritmo agrupa puntos de alta densidad y considera los puntos alejados como ruido o valores atípicos. Es ideal para identificar clusters de formas arbitrarias en grandes conjuntos de datos (Castillo, 2023).

Para profundizar en este análisis, es necesario destacar el papel de los algoritmos de clustering más comunes que son ampliamente utilizados en el ámbito geoespacial por su capacidad para identificar patrones. Estos algoritmos permiten abordar de manera eficiente la complejidad de los datos espaciales, ofreciendo soluciones robustas para el análisis de fenómenos que presentan distribuciones heterogéneas y formas arbitrarias.

El objetivo de este estudio es realizar una revisión sistemática que permita analizar la implementación de algoritmos de clustering en el contexto de datos geoespaciales, destacando la integración de técnicas de inteligencia artificial como herramientas clave para la identificación y análisis de patrones espaciales en Sistemas de Información Geográfica (SIG). A través de este

enfoque, se busca resaltar la importancia de estas tecnologías en la comprensión de fenómenos complejos y su aplicabilidad en diversos ámbitos de estudio.

Metodología

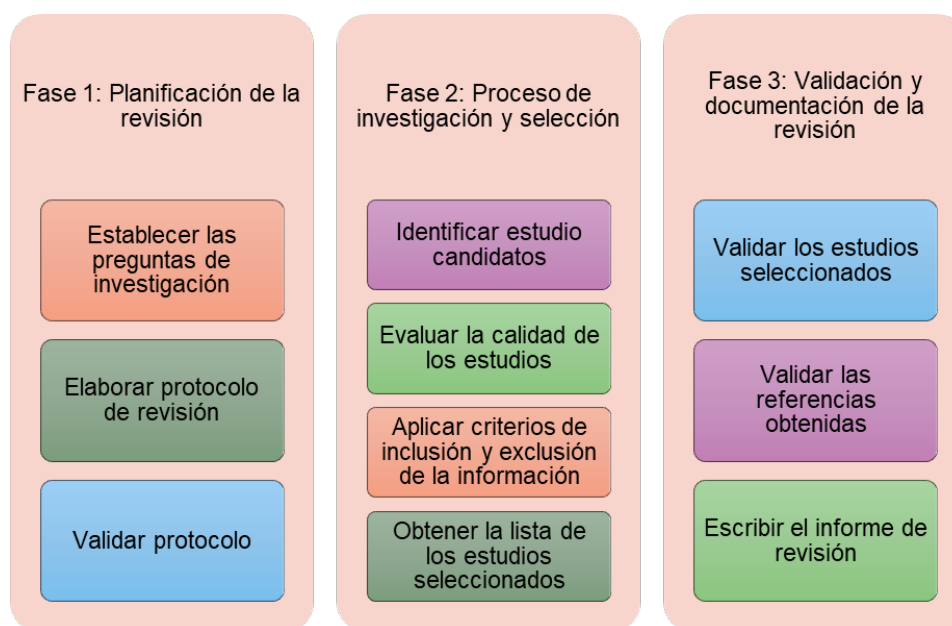
Posteriormente, se procedió a la clasificación de los documentos seleccionados a partir de criterios previamente definidos, tales como el tipo de algoritmo utilizado, el ámbito disciplinario de aplicación y las métricas de evaluación reportadas. Para garantizar consistencia, la información fue extraída mediante un formulario estructurado que recopiló variables bibliográficas, técnicas y metodológicas. La extracción de datos fue realizada por un revisor, y las discrepancias presentadas al consultar los resultados con un segundo revisor se resolvieron mediante discusión conjunta; en caso de persistir, se recurrió a un tercer revisor.

Una vez organizada la información, se llevó a cabo un análisis crítico basado en la comparación sistemática entre estudios. Este análisis consideró similitudes, diferencias, aportaciones metodológicas, limitaciones reportadas y el desempeño de los algoritmos en contextos geoespaciales. Este procedimiento permitió sintetizar los avances más significativos en el campo mediante una síntesis narrativa y tabulada, agrupando los hallazgos por tipo de algoritmo, ámbito de aplicación (por ejemplo, arqueología, sociología, economía o criminología) y métricas de validación utilizadas.

La investigación se basó en la aplicación de métodos históricos, sintéticos y analíticos, que permitieron no solo trazar la evolución de los algoritmos de clustering, sino también identificar su pertinencia en diversas disciplinas. De este modo, se proporcionó una visión integral de las posibilidades y desafíos que presentan estas técnicas en el contexto geoespacial.

Se decidió seguir las pautas metodológicas propuestas por Kitchenham y Brereton (2013), para revisiones sistemáticas en ingeniería de software. Esta guía proporcionó un marco para asegurar una búsqueda profunda, un proceso de selección riguroso y una evaluación crítica fundamentada en criterios reproducibles, lo que permitió establecer un protocolo claro para la revisión documental (ver Figura 1). Además, esta metodología promovió la transparencia del estudio, aspecto fundamental para validar los hallazgos en investigaciones de carácter interdisciplinario como la presente.

Figura 1. Proceso de selección de información de Kitchenham, y Brereton (2013).



Fuente: elaboración propia

Fase 1: planificación de la revisión

Siguiendo esta metodología, inicialmente se definen las preguntas de investigación que son el objetivo principal de la realización de este análisis sistemático. Actualmente se disponen de herramientas que facilitan el análisis de datos de información geográfica, una de ellas son los sistemas de información geográfica.

Los avances en los procesos de análisis de datos y el aprendizaje automático ofrecen oportunidades para abordar cualquier problemática en cualquier escenario que requiera examinar información geoespacial. Sin embargo, aún existen problemas con la identificación y la interpretación de patrones en cuanto a la distribución geoespacial ya que, en muchas ocasiones, la segmentación espacial no es del todo precisa para manejar datos complejos.

Los algoritmos de clustering ofrecen un panorama favorable para el análisis de datos geoespaciales que muestran cierto grado de complejidad. Si estos pueden ser adaptados a un sistema de información geográfica, es posible robustecer los procesos en el análisis de datos geoespaciales. Derivado de lo anterior surgen las preguntas de investigación:

- ¿Cuál es el impacto de las técnicas de la inteligencia artificial en los sistemas de información geográfica según los hallazgos obtenidos a través de una revisión sistemática?
- ¿Cuáles son los algoritmos de clustering más utilizados para el análisis de datos geoespaciales?

Fase 2: proceso de investigación y selección

Se definieron criterios de inclusión y exclusión de las fuentes revisadas para evaluar la calidad de los estudios seleccionados y se sintetizaron los resultados a través de un proceso sistemático y controlado, lo que permitió obtener una visión comprensiva y fiable de la implementación de herramientas de IA en el análisis de datos geográficos o espaciales.

Para realizar una búsqueda de artículos científicos relacionados con las palabras clave “Clustering”, “GIS” y/o Inteligencia Artificial, se utilizó un conjunto de bases de datos académicas, incluyendo ScienceDirect, ResearchGate, SciELO, Springer, Redalyc y Dialnet. La Tabla 1 describe el proceso de búsqueda y los filtros aplicados en cada base de datos para excluir resultados irrelevantes.

Tabla 1. Filtros aplicados en la búsqueda en bases de datos.

Base de datos	Filtros aplicados
ScienceDirect	Palabras clave “Clustering” y “GIS”. Filtros para incluir artículos de investigación (Research articles) que sean de acceso abierto (Open Access) y del área de las ciencias de la computación (Computer Science).
ResearchGate	Palabras clave “Clustering” y “GIS”. de búsqueda para incluir artículos de revista y de acceso abierto.
SciELO	Términos de búsqueda en español, “Clustering” e “Inteligencia Artificial” Los filtros se ajustaron incluir estudios de investigaciones relacionadas con ingenierías.
Springer	Palabras clave “Clustering Algorithms”. Filtros aplicados para incluir el tipo de información (Geographical information system) y áreas (Computer science y Artificial intelligence).
Redalyc	Búsqueda de “Clustering” e “Inteligencia Artificial”, Incluyendo los artículos que pertenecen a las áreas de computación y política de los últimos 2 años (2023 y 2024).
Dialnet	Búsqueda del término “Clustering” e “Inteligencia Artificial”. Aplicación de filtros centrándose en artículos de revista.

Fuente: elaboración propia

En la Tabla 2, se muestra un cuadro con el número de resultados obtenidos en cada base de datos tras aplicar los filtros mencionados:

Tabla 2. Cifras de resultados obtenidos en la búsqueda.

Base de datos	Resultados totales	Resultados tras aplicar filtros
ScienceDirect	1,076	479
ResearchGate	7,671	620
SciELO	3,103	526
Springer	36,688	181
Redalyc	773,628	301

Base de datos	Resultados totales	Resultados tras aplicar filtros
Dialnet	131	46
Total	822,297	2,153

Fuente: elaboración propia

Después de obtener los estudios, se aplicaron algunos criterios de inclusión y exclusión para obtener la revisión sistemática.

Criterios de inclusión

- Estudios que implementan algoritmos de clustering aplicados al análisis de datos geoespaciales.
- Artículos publicados en revistas y conferencias académicas reconocidas, con revisión por pares y publicados en los últimos 10 años.
- Investigaciones que integran técnicas de inteligencia artificial en los sistemas de información geográfica para identificar patrones espaciales.
- Artículos que proporcionan información sobre metodologías, aplicaciones o análisis empíricos en el uso de SIG y clustering en diversas disciplinas.
- Investigaciones que incluyen evaluación de la efectividad y precisión de algoritmos de clustering en la detección de patrones espaciales.

Criterios de exclusión

- Estudios que no utilizan herramientas de inteligencia artificial o se enfocan en otros tipos de análisis sin aplicación espacial.
- Artículos de fuentes no revisadas por pares, como blogs, presentaciones informales o reportes sin metodología académica.
- Investigaciones que no integran técnicas de inteligencia artificial en los sistemas de información geográfica (SIG).
- Estudios que no incluyen el uso de inteligencia artificial ni presentan relación con SIG en sus aplicaciones.
- Estudios escritos en idiomas distintos al inglés o español.
- Documentos que no presentan aplicación práctica o estudios empíricos y se limitan a teorías no aplicadas en el contexto geoespacial

Fase 3: validación y documentación de la revisión

En esta fase se excluyeron las investigaciones localizadas durante diciembre de 2024, manteniendo solo las que resultaron relevantes para el estudio (ver Tabla 3).

Tabla 3. Cantidad de estudios encontrados por base de datos.

Base de Datos	Estudios encontrados
ScienceDirect	3
ResearchGate	6
SciELO	0
Springer	3
Redalyc	1
Dialnet	2

Fuente: elaboración propia

Resultados

A partir de la revisión sistemática realizada, se utilizaron bases de datos académicas reconocidas ScienceDirect, ResearchGate, SciELO, Springer, Redalyc y Dialnet aplicando filtros específicos para garantizar la relevancia y calidad de los estudios seleccionados.

En esta etapa, se recopilaron un total de 2,153 artículos tras aplicar criterios de búsqueda iniciales, los cuales fueron refinados mediante filtros avanzados y criterios de inclusión y exclusión rigurosos. Los estudios seleccionados destacan por implementar algoritmos de clustering en sistemas de información geográfica (SIG) y por integrar técnicas de inteligencia artificial en la identificación de patrones espaciales.

Los resultados incluyen un análisis detallado de los artículos seleccionados, clasificados según su procedencia y la aplicabilidad de sus metodologías (ver Tabla 4). Este enfoque permitió reducir significativamente la cantidad de estudios hasta identificar los más relevantes para el objetivo de esta investigación.

Tabla 4. Estudios seleccionados.

Autor	Año	Título	Revista	Citas
Aksoy	2006	Clustering With GIS: An Attempt to Classify Turkish District Data	XXIII International FIG Congress	15
Abubahia y Cocea	2015	A Clustering Approach for Protecting GIS Vector Data	Advanced Information Systems Engineering	17

Autor	Año	Título	Revista	Citas
Biswas et al.	2020	Geospatial Clustering for Balanced and Proximal Schools.	Proceedings of the AAAI Conference on Artificial Intelligence	14
Fontalvo et al.	2023	Método de clustering e inteligencia artificial para clasificar y proyectar delitos violentos en Colombia	Revista Científica General José María Córdova	1
Folini et al.	2022	Cluster analysis: a comprehensive and versatile qgis plugin for pattern recognition in geospatial data	The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLVIII-4/W1-2022	-
García et al.	2017	Detección automática del nivel de crimen basado en el análisis de puntos calientes en la ciudad de Guayaquil. Dominio de las ciencias	Dominio de las ciencias	-
Laskowski y Tomiło	2023	A new AI-based method for clustering survey responses	Journal of Modern Science	0
Lemenkova	2021	Evaluating Land Cover Types From Landsat TM Using SAGA GIS for Vegetation Mapping Based on ISODATA and K-Means Clustering	Acta Agriculturae Serbica	5
Li et al.	2024	GeoAI for Science and the Science of GeoAI	Journal of Spatial Information Science	3
Mouraz et al.	2022	Combining cluster analysis and GIS maps to characterise building stock: case study in the historical city centre of Viseu, Portugal	Journal of Building Engineering	5
Muhammad	2019	Application of labelled k means clustering for GIS contract automation	Journal of Engineering and Applied Sciences	-

Autor	Año	Título	Revista	Citas
Nieto et al.	2023	Data Analysis y Clustering para el análisis de crímenes de alto impacto en Barranquilla	Universidad del norte	-
Roy y Mandal	2012	A Novel Spatial Fuzzy Clustering using Delaunay Triangulation for Large Scale GIS Data (NSFCDT)	Procedia Technology	-
Schoier y Gregorio	2017	Clustering Algorithms for Spatial Big Data	International Conference on Computational Science and Its Applications	7
Velandia et al.	2018	Prototipo de un modelo clasificador para la toma de decisiones a partir de la percepción actual de la seguridad en Bogotá	TIA	-

Fuente: elaboración propia

La literatura revisada muestra una amplia diversidad de enfoques que combinan técnicas de clustering con Sistemas de Información Geográfica (SIG) para el análisis de datos espaciales. Sin embargo, los estudios no aportan conocimientos de manera homogénea: algunos se centran en el desarrollo o mejora de algoritmos, otros en aplicaciones específicas (como la delincuencia, la gestión territorial o la educación) y un tercer grupo aborda la integración de clustering con herramientas de inteligencia artificial o plataformas SIG. A continuación, se muestran los trabajos seleccionados.

Algoritmos de clustering utilizados en los estudios incluidos

Los estudios identificaron el uso de diversos algoritmos de clustering aplicados a datos geoespaciales. K-means fue empleado en investigaciones como las de Lemenkova (2021); Velandia et al. (2018) y Biswas et al. (2020). También se reportó el uso de variantes del método, como GeoKmeans (Biswas et al., 2020) y Labeled K-means (Muhammad, 2019).

Otros algoritmos identificados incluyen ISODATA (Lemenkova, 2021), técnicas de clustering difuso basadas en triangulación de Delaunay (Roy & Mandal, 2012), DBSCAN, FSDP y cluStering jerárquico aglomerativo (Schoier & Gregorio, 2017; Laskowski & Tomiło, 2023).

Aplicaciones del clustering en SIG

Los estudios aplicaron técnicas de clustering en distintos contextos geoespaciales. Se identificaron aplicaciones en clasificación territorial (Aksoy, 2006; Fontalvo et al., 2023), análisis de cobertura terrestre mediante imágenes satelitales (Lemenkova, 2021), caracterización de edificaciones (Mouraz et al., 2022), análisis de puntos calientes del crimen (García et al., 2017; Nieto et al., 2023) y análisis de percepción de seguridad (Velandia et al., 2018). En el ámbito de administración de datos, se reportó el uso del clustering para protección de datos vectoriales mediante marcas de agua (Abubahia & Cocea, 2015) y para la automatización de contratos GIS (Muhammad, 2019).

Integración de clustering con inteligencia artificial

Algunos estudios combinaron técnicas de clustering con métodos de inteligencia artificial. Fontalvo et al. (2023), integraron clustering con redes neuronales para clasificar delitos violentos. Laskowski y Tomiło (2023), aplicaron Autoencoders Variacionales para agrupar respuestas de encuestas y compararon su desempeño con métodos tradicionales. Muhammad (2019), complementó el clustering con redes neuronales de retropropagación para el seguimiento automatizado de contratos GIS.

Plataformas y herramientas SIG empleadas

Se identificó el uso de plataformas especializadas para la visualización o procesamiento geoespacial. Folini et al. (2022), reportaron un complemento para QGIS orientado al análisis de patrones mediante clustering. Otros estudios emplearon bibliotecas de análisis como Pandas y visualización como Plotly (Nieto et al., 2023), además de herramientas estándar de SIG para representación cartográfica (Mouraz et al., 2022; García et al., 2017).

Métricas de desempeño reportadas

Algunos estudios informaron métricas cuantitativas. García et al. (2017), reportaron un 93% de precisión en la clasificación de niveles de criminalidad, y Fontalvo et al. (2023), una precisión del 97,7% en la clasificación de delitos violentos. Los demás trabajos no proporcionaron métricas comparables o se centraron en la descripción metodológica.

Discusión

Los resultados de la revisión sistemática evidencian que la integración de algoritmos de clustering con Sistemas de Información Geográfica (SIG) constituye una estrategia ampliamente utilizada para el análisis de patrones geoespaciales en diversas áreas de aplicación. Los estudios analizados muestran que los algoritmos más empleados corresponden a enfoques basados en densidad y partición, como K-means, DBSCAN y variantes difusas, debido a su capacidad para

manejar grandes volúmenes de datos espaciales y revelar estructuras subyacentes en la información geográfica.

En particular, la literatura revisada indica que la incorporación de técnicas de inteligencia artificial en entornos SIG ha ampliado las capacidades tradicionales de análisis espacial, permitiendo no solo la identificación de agrupamientos espaciales, sino también el tratamiento de datos heterogéneos y de alta dimensionalidad.

Análisis de Estudios Relevantes

La integración de algoritmos de clustering con sistemas de información geográfica ha revolucionado el análisis de datos geoespaciales, proporcionando herramientas novedosas que permiten identificar y comprender patrones complejos en diversas disciplinas. Los SIG permiten la visualización, análisis y manipulación de datos espaciales, mientras que los algoritmos de clustering agrupan datos similares para descubrir relaciones y tendencias ocultas. Esta sinergia facilita la exploración de grandes volúmenes de información geográfica, permitiendo a investigadores y profesionales en campos como la arqueología, la sociología, la economía y la criminología identificar patrones espaciales significativos y tomar decisiones informadas. A medida que la disponibilidad de datos geoespaciales sigue creciendo, la combinación de algoritmos de clustering y SIG se convierte en una herramienta esencial para abordar desafíos contemporáneos, optimizando la planificación y gestión en diversas áreas de estudio como las siguientes:

Roy, y Mandal en 2012 desarrollaron un método de clustering difuso basado en la triangulación de Delaunay para datos SIG a gran escala. Su enfoque integra una primera fase de agrupamiento espacial basado en la proximidad, seguido de una refinación mediante técnicas difusas, lo que permite mejorar la precisión en la identificación de patrones espaciales de interés.

Schoier y Gregorio (2017), evaluaron la eficiencia de tres algoritmos de clustering: DBSCAN, FSDP y K-means, en el análisis de grandes volúmenes de datos espaciales. Su estudio destacó la modificación del algoritmo FSDP para mejorar su eficiencia en bases de datos grandes, lo cual es relevante en el contexto de la delincuencia, donde los datos pueden ser extensos y heterogéneos.

García, Villota y Litardo en el año 2017 llevaron a cabo un análisis de puntos calientes en Guayaquil, clasificando niveles de criminalidad en diferentes zonas geográficas con una precisión del 93%. Estos estudios subrayan la relevancia de las técnicas de clustering para la identificación de áreas críticas, permitiendo a las autoridades orientar mejor sus recursos para la prevención del crimen.

Abubahia y Cocea (2015), presentan una metodología para proteger datos de sistemas de información geográfica mediante la incrustación de marcas de agua digitales en los datos vectoriales. Utilizan técnicas de clustering para aumentar la robustez de la protección contra modificaciones indeseadas como la simplificación y la interpolación de los datos. Esta metodología es relevante para el análisis de patrones espaciales de delincuencia, ya que subraya la importancia

de mantener la integridad de los datos geoespaciales al tiempo que se utilizan para identificar patrones criminales.

La combinación de técnicas de clustering con SIG es comúnmente utilizada para visualizar y analizar datos espaciales complejos. Mouraz et al. (2022), emplearon mapas basados en sistemas de información geográfica para visualizar los resultados del análisis de clustering sobre defectos en edificaciones, lo que facilitó la identificación de patrones relacionados con la distribución espacial de los defectos. Este enfoque puede ser extrapolado al análisis de la delincuencia, donde los mapas SIG permiten una representación visual intuitiva de las zonas de alta criminalidad.

Biswas, Chen, Sistrunk, Muthiah, Chen Z., Self, Lu, y Ramakrishnan en el 2020 propusieron el algoritmo GeoKmeans para la redistrictación escolar, enfrentando retos similares a los del clustering de delitos, como la contigüidad geográfica y la equilibración de restricciones específicas del problema. Su metodología, basada en geometrías poligonales y criterios de proximidad, puede ser adaptada para el diseño de estrategias de intervención en áreas con alta incidencia delictiva.

En el año 2023, Fontalvo, Vega, y Mejía, combinaron técnicas de clustering con redes neuronales para la clasificación y pronóstico de delitos violentos en Colombia, alcanzando una precisión del 97.7% en la clasificación de los tipos de delitos. Esta integración de clustering con inteligencia artificial ofrece un enfoque poderoso para el análisis predictivo, permitiendo no solo identificar patrones existentes sino también predecir tendencias futuras.

Si bien existen múltiples enfoques para la identificación de patrones espaciales de criminalidad, la falta de universalidad en las técnicas de clustering (señalado por Aksoy en 2006) implica que cada contexto requiere un ajuste específico de las metodologías. La capacidad de combinar diversos algoritmos y adaptar los parámetros a los datos disponibles sigue siendo un desafío central.

Además, la integración de clustering con herramientas modernas de SIG y plataformas abiertas, como el plugin de análisis de clusters desarrollado por Folini, Lenzi, y Biraghi, en 2022 para QGIS, representa un avance significativo hacia la democratización y accesibilidad de estas técnicas para diferentes usuarios, incluyendo autoridades de seguridad.

En el 2023 Laskowski, y Tomiło, introducen un método basado en inteligencia artificial (AI) para la agrupación de respuestas de encuestas utilizando un Autoencoder Variacional (VAE). Al comparar este enfoque con métodos tradicionales como K-means y el clustering jerárquico aglomerativo, demuestran una mejora significativa en la captura de relaciones complejas y patrones en los datos. Este estudio destaca la aplicabilidad de técnicas avanzadas de clustering para analizar datos complejos, como los patrones de criminalidad, y su resistencia al ruido en los datos.

Nieto et al. (2023), desarrollan un enfoque para el análisis de crímenes de alto impacto en Barranquilla utilizando bibliotecas de Python como Pandas y Plotly. Estas herramientas permiten la visualización de datos a través de gráficos y mapas de calor, facilitando la identificación de puntos críticos de violencia. El uso de clustering ayuda a identificar patrones espaciales de criminalidad, proporcionando una base para la implementación de medidas preventivas.

Velandia, Guataquirá y Prada en 2018 proponen un modelo clasificador utilizando minería de datos y el algoritmo K-means para analizar la criminalidad en Bogotá. A través del clustering, identifican patrones no evidentes a simple vista y evalúan factores relevantes para la prevención del delito. El estudio enfatiza la utilidad del clustering para la toma de decisiones basadas en datos y la formulación de políticas de seguridad.

Lemenkova en el 2021 explora la clasificación no supervisada mediante ISODATA y K-means en imágenes Landsat para el mapeo de la cobertura terrestre en Islandia. Los métodos demostraron ser efectivos para la clasificación de tipos de cobertura del suelo, lo que ilustra el potencial del clustering en el procesamiento de datos de sensores remotos y en la cartografía. Este enfoque puede extenderse al análisis de patrones espaciales de criminalidad, integrando datos ambientales y geográficos para un análisis más completo.

Li et al. (2024), analizaron las tendencias actuales en la investigación de GeoAI y abordaron los avances más innovadores en este campo, destacando su papel en la aceleración de las ciencias ambientales y sociales. Abordaron los esfuerzos para mejorar la predictibilidad de los modelos de GeoAI, así como las investigaciones recientes dirigidas a aumentar su explicabilidad y reproducibilidad, con el fin de garantizar hallazgos geoespaciales confiables.

Muhammad en 2019 propuso una solución para automatizar la gestión técnica de contratos de Sistemas de Información Geográfica (GIS) mediante la aplicación de Labeled K Means Clustering. Exploraron la limitación actual del seguimiento manual de contratos GIS, que carece de indicadores automáticos y mecanismos centralizados para evaluar el cumplimiento de las partes involucradas. La propuesta incluye convertir contratos manuales en electrónicos a través de un proceso de agrupamiento basado en lexemas y almacenarlos en una base de datos centralizada. Además, utiliza una Red Neuronal de Retropropagación (Back Propagation Neural Network) para entrenar el sistema en indicadores de desempeño, acompañado de una interfaz gráfica asistida para facilitar la regulación y el seguimiento de los contratos.

A lo largo del análisis, se evidenció que estos algoritmos no solo facilitan la segmentación y el análisis de datos, sino que también ofrecen herramientas valiosas para la comprensión de fenómenos complejos no solo de criminalidad sino de diversas disciplinas cartográficas.

Los estudios evidencian que la visualización geoespacial proporcionada por los SIG desempeña un papel clave en la interpretación de los resultados del clustering, ya que facilita la identificación de zonas críticas, tendencias espaciales y relaciones entre variables geográficas. Esta característica es especialmente relevante en contextos de seguridad y criminología, donde la representación espacial de los datos contribuye a una mejor comprensión de la distribución del delito y al apoyo en la toma de decisiones estratégicas.

La revisión también pone de manifiesto una limitada universalidad en el uso de algoritmos de clustering, ya que la selección del método y la configuración de sus parámetros dependen en

gran medida del contexto, el tipo de datos y los objetivos del análisis. Esta falta de estandarización coincide con lo reportado en la literatura, donde se señala que no existe un algoritmo óptimo para todos los escenarios geoespaciales, lo que representa un desafío recurrente en este tipo de estudios.

La capacidad de los SIG para visualizar y analizar datos geoespaciales, combinada con los algoritmos de clustering, permite a los profesionales en materia de seguridad a descubrir tendencias y relaciones que de otro modo podrían pasar desapercibidas. Esta integración potencia la investigación académica y podría tener implicaciones prácticas en la toma de decisiones y la planificación estratégica. A medida que los datos geoespaciales continúan proliferando, la utilización de algoritmos de clustering en conjunto con los SIG se presenta como una vía prometedora para el avance del conocimiento y la innovación en el análisis de datos en múltiples áreas de estudio.

Conclusión

La revisión sistemática realizada permitió identificar y analizar catorce estudios que aplican algoritmos de clustering en el contexto de los sistemas de información geográfica. Con base en los resultados, se observa que el uso de técnicas de clustering constituye un componente recurrente en aplicaciones geoespaciales, particularmente en tareas de clasificación, segmentación y detección de patrones espaciales. Los algoritmos más empleados fueron K-means y DBSCAN, junto con variantes especializadas como GeoKMeans, Labeled K-means y enfoques basados en autoencoders variacionales. Estos hallazgos responden al objetivo de describir la implementación de técnicas de clustering en SIG y evidencian una tendencia hacia la incorporación de métodos de inteligencia artificial en los estudios más recientes.

Los SIG tienen influencia en los estudios incluidos, muestran que su integración se utiliza principalmente para mejorar la precisión y la capacidad de modelado, especialmente en aplicaciones como la clasificación de delitos, la protección de datos geoespaciales, la automatización de procesos GIS y el análisis de imágenes satelitales. La presencia de GeoAI en trabajos recientes presenta un avance progresivo hacia modelos más complejos y explicables, lo que contribuye al fortalecimiento de las ciencias espaciales.

En conjunto, los resultados evidencian que la implementación de algoritmos de clustering en SIG se encuentra ampliamente distribuida entre diversas disciplinas, lo que señala su utilidad como herramienta analítica para el estudio de patrones espaciales. Asimismo, los estudios aplicados a criminalidad demuestran que el clustering es una técnica utilizada con frecuencia en la identificación de concentraciones geográficas de incidentes y en el apoyo a procesos de análisis territorial.

Es importante resaltar que esta revisión presenta limitaciones. La heterogeneidad metodológica entre los estudios dificulta la comparación directa de sus resultados. Además, la mayoría de las investigaciones no presentan métricas estandarizadas de evaluación, lo que

restringe la posibilidad de establecer conclusiones cuantitativas sobre el desempeño relativo de los algoritmos. Finalmente, la revisión se limitó a artículos publicados en inglés y español, lo cual puede excluir aportes relevantes en otros idiomas.

La revisión sistemática cumple el objetivo de caracterizar el uso de algoritmos de clustering en SIG y de identificar el papel creciente de la inteligencia artificial en este campo. Los patrones observados en los estudios permiten afirmar que el clustering constituye una herramienta ampliamente implementada en el análisis geoespacial y que su integración con enfoques de IA continúa en expansión, consolidándose como un componente clave en el desarrollo de aplicaciones geográficas contemporáneas.

Referencias

- Abubahia, A., & Cocea, M. (2015). A clustering approach for protecting GIS vector data. In J. Zdravkovic, M. Kirikova, & P. Johannesson, (eds.). *Advanced Information Systems Engineering* (pp. 133–147). Springer. https://doi.org/10.1007/978-3-319-19069-3_9
- Aksoy, E. (2006, octubre 8-13). *Clustering with GIS: An attempt to classify Turkish district data* [Conferencia]. XXIII FIG International Congress, Munich, Alemania.
- Amat, J. (2017). *Clustering y heatmaps: Aprendizaje no supervisado*. https://rpubs.com/Joaquin_AR/310090
- Biswas, S., Chen, F., Sistrunk, A., Muthiah, S., Chen, Z., Self, N., Lu, C.-T., & Ramakrishnan, N. (2020). Geospatial clustering for balanced and proximal schools. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(9), 13358–13365. <https://doi.org/10.1609/aaai.v34i09.7058>
- Castillo, G. (2023, 16 de febrero). *Aprendizaje no supervisado: ¿Qué es y cómo funciona?* NeuroK. <https://neurok.es/aprendizaje-no-supervisado-que-es/>
- Folini, A., Lenzi, E., & Biraghi, C. A. (2022). Cluster analysis: A comprehensive and versatile QGIS plugin for pattern recognition in geospatial data. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVIII-4/W1-2022, 151–157. <https://doi.org/10.5194/isprs-archives-XLVIII-4-W1-2022-151-2022>
- Fontalvo, H. T., Vega, H. M., & Mejía, Z. F. (2023). Método de clustering e inteligencia artificial para clasificar y proyectar delitos violentos en Colombia. *Revista Científica General José María Córdova*, 21(42), 551–572. <https://doi.org/10.21830/19006586.1117>
- García, P., Villota, O. W., & Litardo, U. J. (2017). Detección automática del nivel de crimen basado en el análisis de puntos calientes en la ciudad de Guayaquil. *Dominio de las Ciencias*, 3(2), 367–379.
- Kitchenham, B., & Brereton, P. (2013). A systematic review of systematic review process research in software engineering. *Information and Software Technology*, 55(12), 2049–2075. <https://doi.org/10.1016/j.infsof.2013.07.010>
- Laskowski, J. F., & Tomiło, P. (2023). A new AI-based method for clustering survey responses. *Journal of Modern Science*, 54(5), 355–377. <https://doi.org/10.13166/jms/176171>

- Lemenkova, P. (2021). Evaluating land cover types from Landsat TM using SAGA GIS for vegetation mapping based on ISODATA and k-means clustering. *Acta Agriculturae Serbica*, 26(52), 159–165. <https://doi.org/10.5937/AASer2152159L>
- Li, W., Arundel, S., Gao, S., Goodchild, M., Hu, Y., Wang, S., & Zipf, A. (2024). GeoAI for science and the science of GeoAI. *Journal of Spatial Information Science*, 29, 1–17. <https://doi.org/10.5311/JOSIS.2024.29.349>
- Mohammad, N. (2023). *A computational theory and semi-supervised algorithm for clustering*. arXiv. <https://doi.org/10.48550/arXiv.2306.06974>
- Mouraz, C. P., Almeida, R. M. S. F., & Mendes, S. J. (2022). Combining cluster analysis and GIS maps to characterise building stock: Case study in the historical city centre of Viseu, Portugal. *Journal of Building Engineering*, 58. <https://doi.org/10.1016/j.jobbe.2022.104949>
- Muhammad, S. (2019). Application of labelled k means clustering for GIS contract automation. *Journal of Engineering and Applied Sciences*, 38(1), 25–34.
- Nieto, W., Plata, D., Torres, R., & Solano, E. (2023). *Data analysis y clustering para el análisis de crímenes de alto impacto en Barranquilla* [Tesis de grado, Universidad del Norte].
- Pedada, S. (2023, 07 de junio). ¿Qué es el análisis geoespacial? El plan antes del plan real. Mind the Graph. <https://mindthegraph.com/blog/es/analisis-geoespacial/>
- Ramírez, L. (2023, 11 de abril). *Algoritmo k-means: ¿Qué es y cómo funciona?* IEBS. <https://n9.cl/8zbcu7>
- Roy, P., & Mandal, J. K. (2012). A novel spatial fuzzy clustering using Delaunay triangulation for large scale GIS data (NSFCDT). *Procedia Technology*, 6, 452–459. <https://doi.org/10.1016/j.protcy.2012.10.054>
- Sancho, C. F. (2023). *Algoritmos de clustering*. MatematIA. <https://n9.cl/2f0tp4>
- Schoier, G., & Gregorio, C. (2017). Clustering algorithms for spatial big data. In O. Gervasi, B. Murgante, S. Misra, G. Borruso, C. M. Torre, A. M. A. C. Rocha, D. Taniar, B. O. Apduhan, E. Stankova, & A. Cuzzocrea, (eds.). *Computational Science and Its Applications – ICCSA 2017* (pp. 571–583). Springer. https://doi.org/10.1007/978-3-319-62401-3_41
- Velandia, R., Guataquira, A., & Prada, J. (2018). Prototipo de un modelo clasificador para la toma de decisiones a partir de la percepción actual de la seguridad en Bogotá. *Tecnología, Innovación y Aplicación*, 6(2), 18–24.
- Vichi, M., Cavicchia, C., & Groenen, P. J. F. (2022). Hierarchical means clustering. *Journal of Classification*, 39(3), 553–577. <https://doi.org/10.1007/s00357-022-09419-7>

Autores

José Roman Castro San Agustín. Maestro en Ciencias de la computación y actualmente estudiante del Doctorado en Ciencias de la computación en la Universidad Autónoma del Estado de México Campus Valle de Chalco.

Marco Alberto Mendoza Pérez. Profesor de Tiempo Completo, investigador, docente de Ingeniería en Computación, Maestría y Doctorado en Ciencias de la Computación, y responsable del Laboratorio de Electrónica en la Universidad Autónoma del Estado de México (UAEMex)/Centro Universitario Valle de Chalco.

Anabelem Soberanes Martín. Coordinadora del Doctorado en Ciencias de la Computación y profesora de tiempo completo, impartiendo clases en las licenciaturas de Ingeniería en Computación e Informática Administrativa, así como en los programas de maestría y doctorado en Ciencias de la Computación.

Declaración

Conflicto de interés

No tenemos ningún conflicto de interés que declarar.

Financiamiento

Sin ayuda financiera de partes externas a este artículo.

Agradecimientos

Consejo Nacional de Humanidades, Ciencia y Tecnologías (SECIHTI) por su apoyo con la beca (CVU: 1139656) en este proyecto.

Nota

El artículo es original y no ha sido publicado previamente.