DIOTIMA
REVISTA CIENTÍFICA DE ESTUDIOS TRANSDISCIPLINARIA
Vol. 10 No. 30 Septiembre-diciembre (2025)
ISSN: 2448-5497

Evaluación de comportamiento de algoritmos de Clustering para la identificación de patrones de delincuencia

Behavioral evaluation of Clustering algorithms for identifying crime patterns

José Román Castro San Agustín[1]

jcastros248@alumno.uaemex.mx

ORCID: 0000-0001-6383-6700

Marco Alberto Mendoza Pérez

mamendozap@uaemex.mx

ORCID: 0000-0003-4911-4757

Doricela Gutiérrez Cruz

dgutierrezcr@uaemex.mx

ORCID: 0000-0003-2843-3273

Cristina Juárez Landín

cjuarezl@uaemex.mx

ORCID: 0000-0002-0988-3060

[1] Autores del Doctorado en Ciencias de la Computación, Centro Universitario UAEM Valle de Chalco y Centro Universitario UAEM Nezahualcóyotl, Universidad Autónoma del Estado de México.

Resumen

El presente trabajo de investigación tuvo como objetivo evaluar las zonas de alto riesgo en la Ciudad de México mediante la aplicación de técnicas de aprendizaje no supervisado, específicamente utilizando algoritmos de clustering. La investigación se centró en la categorización y agrupación de datos espaciales relacionados con incidentes delictivos, con el fin de identificar patrones útiles para mejorar la toma de decisiones en materia de seguridad pública. Para procesar y analizar los datos, se implementaron los algoritmos de clustering DBSCAN (Density-Based Spatial Clustering of Applications with Noise), K-means, clustering jerárquico y Mean Shift, seleccionados por su capacidad de manejar datos espaciales y revelar agrupamientos con características distintas. Los datos utilizados provinieron de fuentes oficiales del Gobierno de la Ciudad de México e incluyeron coordenadas geográficas de delitos registrados. Cada algoritmo fue evaluado con base en su desempeño para identificar zonas de concentración delictiva, considerando la cohesión de los grupos formados, su capacidad para manejar ruido y su utilidad en la visualización de patrones. Como resultado, se generaron mapas que agruparon las áreas de acuerdo a la densidad de los datos, permitiendo identificar zonas prioritarias que no eran evidentes en un análisis superficial. La aplicación de estos algoritmos permitió detectar patrones espaciales de delincuencia con mayor precisión y eficiencia, facilitando la segmentación del territorio en áreas de atención estratégica. Estos hallazgos pueden contribuir significativamente a la planeación y asignación de recursos para la prevención del delito en contextos urbanos.

Vol. 10 No. 30 Septiembre-diciembre (2025)
ISSN: 2448-5497

Palabras clave: Aprendizaje no supervisado, Clustering, Datos espaciales, Seguridad pública, Zonas de alto riesgo.

Abstract

This research aimed to evaluate high-risk areas in Mexico City through the application of unsupervised learning techniques, specifically using clustering algorithms. The study focused on the categorization and grouping of spatial data related to criminal incidents, with the objective of identifying patterns that could support decision-making in public security. To process and analyze the data, four clustering algorithms were implemented: DBSCAN (Density-Based Spatial Clustering of Applications with Noise), K-means, hierarchical clustering, and Mean Shift. These algorithms were selected based on their ability to handle spatial data and uncover groupings with distinct characteristics. The data used in this study were obtained from official sources provided by the Government of Mexico City and included geographic coordinates of recorded crimes. Each algorithm was evaluated based on its performance in identifying crime concentration zones, taking into account cluster cohesion, noise handling capacity, and effectiveness in pattern visualization. As a result, thematic maps were generated to classify areas according to their level of risk, enabling the identification of priority zones that were not evident through superficial analysis. The application of these algorithms enabled the detection of spatial crime patterns with greater accuracy and efficiency, facilitating the segmentation of urban territory into areas for strategic intervention. These findings can significantly contribute to planning and resource allocation for crime prevention in urban contexts.

***Keywords*:** Unsupervised learning, Clustering, Spatial data, Public security, High-risk areas.

Fecha de envío: 20/05/2025

Fecha de aprobación: 18/07/2025

Fecha de publicación: 01/09/2025

Introducción

La gestión del orden en las actividades urbanas plantea desafíos constantes tanto para las autoridades como para la sociedad. En este contexto, comprender los patrones delictivos e identificar áreas de alto riesgo se ha convertido en un aspecto clave para planificar estrategias preventivas orientadas a la seguridad pública.

En la Ciudad de México, la problemática de la criminalidad es importante a considerar. El boletín estadístico de la incidencia delictiva correspondiente a marzo de 2025 reportó un total de 18,552 carpetas de investigación por delitos del fuero común, destacando 1,134 casos de homicidios dolosos y culposos, 116 delitos contra la libertad personal y 3,151 delitos contra la familia, lo que incluye casos de violencia familiar. Las alcaldías de Cuauhtémoc, Iztapalapa y Gustavo A. Madero registraron la mayor concentración delictiva con 2,817, 2,445 y 1,764 delitos respectivamente, evidenciando zonas críticas que requieren atención prioritaria (Unidad de Estadística y Transparencia de la Ciudad de México, 2025).

Vol. 10 No. 30 Septiembre-diciembre (2025)
ISSN: 2448-5497

Frente a esta situación, el análisis de datos espaciales se presenta como una herramienta para examinar la distribución y dinámica de los delitos en entornos urbanos. Particularmente, las técnicas de aprendizaje automático, y en especial el aprendizaje no supervisado, permiten explorar grandes volúmenes de datos sin necesidad de etiquetas predefinidas, identificando estructuras y patrones ocultos.

Entre estas técnicas, el clustering se ha consolidado como una estrategia eficaz para agrupar datos con características similares en conglomerados o clusters. Esta metodología ha demostrado su utilidad en áreas como la ciberseguridad, la biología molecular y la detección de fraudes, y tiene un enorme potencial en el análisis de seguridad urbana. Los algoritmos de clustering pueden clasificarse en varios tipos por partición, por densidad, jerárquicos, entre otros, cada uno con ventajas específicas según la naturaleza de los datos (Barreno, Bregón, y Martínez, 2021; Mohammad, 2023).

En este trabajo se aplicaron y evaluaron distintos algoritmos de clustering sobre datos georreferenciados de delitos ocurridos en la Ciudad de México, con el objetivo de identificar zonas de alto riesgo y detectar patrones espaciales de criminalidad que puedan orientar la toma de decisiones en materia de seguridad pública. Los algoritmos utilizados fueron K-means, DBSCAN, clustering jerárquico y Mean Shift, seleccionados por su capacidad de adaptación a diferentes estructuras espaciales.

A través de la implementación de estas técnicas, se buscó generar patrones geoespaciales que permitieran visualizar las zonas con mayor concentración delictiva, aportando una base analítica para la intervención estratégica y la asignación eficiente de recursos en políticas de prevención del delito.

Trabajos relacionados

En la actualidad, el desarrollo de tecnologías basadas en algoritmos de clustering han tomado relevancia en distintos enfoques. En 2021, Vera desarrolló un algoritmo evolutivo para la perfilación geográfica criminal, combinando técnicas genéticas y evolutivas para predecir eventos delictivos basándose en patrones temporales y espaciales. En 2022, Carrasco y Moyotl usaron el método OPTICS para detectar "hot spots delictivos" en la Ciudad de México, identificando áreas de alta incidencia que no corresponden con divisiones territoriales comunes. En 2023, Fontalvo, Vega y Mejía aplicaron una red neuronal para clasificar y predecir delitos violentos en Colombia con un 97.7% de precisión, diferenciando regiones según su nivel de impacto delictivo.

Este articulo pretende evaluar y mostrar los patrones generados en las zonas donde se tienen registrados los delitos en la Ciudad de México mediante la aplicación de algoritmos de clustering. Este enfoque tecnológico propone una evolución significativa en los sistemas actuales de seguridad urbana, integrando soluciones avanzadas que potencian la gestión de datos y el análisis espacial para enfrentar los desafíos contemporáneos en materia de seguridad.

Método

La presente investigación se desarrolló bajo un enfoque cuantitativo y no experimental, con un diseño exploratorio-comparativo. El objetivo principal fue evaluar el comportamiento de

Vol. 10 No. 30 Septiembre-diciembre (2025)
ISSN: 2448-5497

distintos algoritmos de clustering aplicados a datos geoespaciales de incidentes delictivos, con el fin de identificar zonas de alto riesgo en la Ciudad de México.

Se obtuvieron bases de datos oficiales del Gobierno de la Ciudad de México que registran incidentes delictivos georreferenciados. Estos datos fueron estructurados para su correcto procesamiento, asegurando la calidad y consistencia de la información espacial. Este conjunto de datos contiene información detallada sobre cada incidente, incluyendo la ubicación geográfica (coordenadas de latitud y longitud), el tipo de delito (como robo, asalto o vandalismo), así como la fecha y hora en que ocurrió. Esta base contenía originalmente 239,685 registros georreferenciados, que abarcaban una amplia gama de tipos de delitos y distintos periodos temporales.

Con el objetivo de realizar un análisis más focalizado y reducir el sesgo generado por la heterogeneidad del conjunto de datos, se aplicaron una serie de filtros de depuración y selección. En primer lugar, se eliminó todo registro que no contara con coordenadas geográficas válidas (latitud y longitud), así como aquellos con información incompleta en campos clave como el tipo de delito o la fecha del incidente.

Posteriormente, se aplicó un filtro temático y temporal, restringiendo el análisis exclusivamente a los casos de “robo a transeúntes en vía pública” ocurridos en el mes de diciembre de 2022. La elección de este tipo específico de delito responde a su alta frecuencia y relevancia social, ya que impacta directamente en la percepción de inseguridad de la ciudadanía. Asimismo, el periodo de análisis fue acotado intencionalmente a un solo mes para facilitar la comparación entre algoritmos de clustering bajo condiciones homogéneas y para evitar distorsiones provocadas por la variabilidad temporal.

Luego de aplicar estos filtros, el conjunto final quedó compuesto por 862 registros, todos ellos con ubicación geográfica precisa, lo cual permitió realizar un análisis espacial robusto y enfocado. Esta reducción en el volumen de datos fue necesaria para garantizar una mayor consistencia en el análisis y una evaluación controlada de los algoritmos, sin perder representatividad del fenómeno delictivo elegido.

Se aplicaron cuatro algoritmos de agrupamiento no supervisado: K-means, DBSCAN, Mean Shift y clustering jerárquico. Cada uno fue ejecutado utilizando el mismo conjunto de datos georreferenciados con el fin de comparar su rendimiento y efectividad en la detección de patrones espaciales de criminalidad.

DBSCAN

DBSCAN (Density-Based Spatial Clustering of Applications with Noise) es un algoritmo de clustering no paramétrico basado en densidades. Su objetivo principal es agrupar puntos de datos en regiones de alta densidad y marcar como *outliers* aquellos puntos que no se ajustan a ninguna de estas regiones densas (Sancho, 2023).

El algoritmo funciona utilizando dos parámetros clave que determinan la densidad mínima necesaria para formar un cluster:

- ε (épsilon): Define el radio de la vecindad que se examina alrededor de cada punto.
- MinPts: Establece el número mínimo de puntos que deben encontrarse dentro del radio ε para considerar que se ha alcanzado una densidad suficiente para formar un cluster.

Vol. 10 No. 30 Septiembre-diciembre (2025)
ISSN: 2448-5497

El algoritmo DBSCAN sigue estos pasos:

1. Para cada punto P en el conjunto de datos D que no ha sido visitado:
   - Se marca como visitado y se calcula su vecindad N dentro de un radio ε.
   - Si el tamaño de N es menor que MinPts, el punto se clasifica como *outlier*.
   - Si el tamaño de N es mayor o igual a MinPts, se crea un cluster que incluye a P y los puntos dentro de N.
2. El proceso de expansión de clusters comienza desde el punto P y se repite para cada punto en la vecindad N. Si un punto dentro de esta vecindad tiene suficientes puntos vecinos (es decir, supera el umbral de densidad), su vecindad también se incorpora al cluster.
3. El proceso se repite hasta que todos los puntos hayan sido visitados y agrupados, o identificados como *outliers*.

El clustering ha sido utilizado ampliamente en la identificación de patrones de criminalidad debido a su capacidad para procesar grandes cantidades de datos y descubrir agrupaciones que no son inmediatamente visibles. Una de sus aplicaciones más comunes es la identificación de “hotspots” o puntos calientes, que son áreas con alta concentración de delitos. Estas detecciones permiten a las fuerzas de seguridad focalizar recursos y estrategias en áreas que realmente lo necesitan, mejorando la eficacia de la vigilancia y la prevención.

En este algoritmo los parámetros principales que afectan el resultado son eps y min_samples. Se eligió un valor de eps = 0.01 (aproximadamente 1.1 km en distancia

geográfica), considerando que este radio permitía captar agrupamientos con una proximidad espacial realista en un entorno urbano como la Ciudad de México. Este valor fue definido tras visualizar diferentes pruebas de densidad y observar que radios mayores unificaban zonas distintas, perdiendo precisión, mientras que radios menores generaban demasiados puntos considerados ruido. El valor de min_samples = 10 se estableció con base en una estimación práctica: se consideró que un mínimo de diez incidentes dentro de un mismo radio era un umbral suficientemente significativo para identificar una concentración delictiva relevante, sin fragmentar excesivamente los datos.

Esta configuración de parámetros permite que *DBSCAN* detecte clusters de formas arbitrarias y excluya puntos dispersos o ruido, lo que es útil para datos espaciales. Sin embargo, una elección inadecuada de eps puede llevar a resultados ineficaces, como detectar demasiados o muy pocos clusters dependiendo de la distribución y densidad de los datos.

K-Means

K-means es un algoritmo de agrupamiento ampliamente utilizado que organiza los datos en un número predeterminado de grupos, representado por K. El objetivo principal es minimizar la variación dentro de los clusters, lo que se logra reduciendo la suma de los cuadrados de las distancias entre cada punto de datos y el centroide de su cluster correspondiente.

El algoritmo funciona iterativamente:

1. Se inicializan K centroides, ya sea de manera aleatoria o mediante alguna estrategia.

Vol. 10 No. 30 Septiembre-diciembre (2025)
ISSN: 2448-5497

2. Cada punto de datos se asigna al cluster cuyo centroide está más cerca, según la distancia euclidiana.
3. Los centroides de los clusters se recalculan en función de los puntos asignados.
4. Este proceso se repite hasta que los centroides no cambien significativamente o se alcance un criterio de convergencia.

El valor óptimo de K, que permite una separación más clara de los datos, no está definido de antemano y depende de las características de los datos. Matemáticamente, el objetivo es minimizar la función de costo definida como (Sancho, 2023):

$$\sum_i \sum_j d(x_{ij}, c_i)^2 \qquad \text{Ec. (1)}$$

En el algoritmo *K-means*, los parámetros clave son el número de clusters a generar, y los centroides que representan los puntos centrales de cada cluster. En este caso, se fijó el número de clusters en n_clusters = 8, lo cual se basó en pruebas exploratorias apoyadas en el método del codo (elbow method), donde se observó una reducción marginal en la inercia a partir de ese punto. Este valor también permitió una comparación directa con los otros algoritmos evaluados, asegurando consistencia analítica. Aunque K-means requiere definir el número de grupos desde el inicio, este valor fue validado con visualización de los resultados y su correspondencia con zonas urbanas claramente diferenciadas.

Este algoritmo es sensible a la elección de los centroides iniciales, lo que puede llevar a resultados subóptimos si no se seleccionan adecuadamente. Sin embargo, su simplicidad y

eficiencia lo hacen ampliamente utilizado para tareas de clasificación y agrupación de datos geoespaciales.

Mean shift

Mean Shift es un algoritmo iterativo de agrupamiento que tiene como objetivo identificar áreas densas de puntos de datos en un espacio multidimensional. A diferencia de otros algoritmos como K-Means, Mean Shift no requiere definir el número de clusters de antemano. Su enfoque principal es localizar los puntos centrales de cada cluster mediante el cálculo de centroides.

Este algoritmo se basa en el concepto de Kernel Density Estimation (KDE), que asume que los datos provienen de una distribución de probabilidad desconocida. Para estimar esta distribución, KDE utiliza funciones llamadas *kernels* que se colocan en cada punto para representar una ponderación en su vecindad. Mean Shift itera desplazando los centroides hacia las áreas de mayor densidad de puntos hasta que convergen, definiendo así los clusters en función de estas zonas densas (Sancho, 2023).

Este algoritmo es útil en aplicaciones donde los clusters tienen formas arbitrarias o no se conoce su número exacto. Los pasos que describen el algoritmo son los siguientes:

1. Inicialización: El proceso comienza con una ventana deslizante de forma circular, centrada en un punto aleatorio C, con un radio r que actúa como núcleo.

Vol. 10 No. 30 Septiembre-diciembre (2025)
ISSN: 2448-5497

2. Desplazamiento iterativo: En cada iteración, la ventana se mueve hacia zonas de mayor densidad. Esto se logra ajustando el centro C hacia la media de los puntos contenidos dentro de la ventana, ya que la densidad es proporcional al número de puntos presentes en ella.
3. Actualización del centroide: Al actualizar el centro C, la ventana se desplaza gradualmente hacia regiones más densas, acercándose al punto de mayor concentración.
4. Convergencia: El movimiento continúa hasta que no sea posible encontrar una dirección que permita incluir más puntos en el núcleo. En este momento, la ventana se considera convergida.
5. Repetición para múltiples ventanas: Este procedimiento se repite para diferentes ventanas iniciales hasta que todos los puntos queden cubiertos. Si varias ventanas se solapan, se retiene aquella que incluye el mayor número de puntos.
6. Agrupamiento final: Una vez que todas las ventanas han alcanzado su convergencia, los puntos se asignan a un cluster según la ventana a la que pertenecen.

En el algoritmo Mean Shift, los parámetros clave son el *bandwidth* (ancho de banda) y la función de estimación de densidad que agrupa los puntos en torno a los modos de la distribución de los datos. El ancho de banda se estima automáticamente en base al parámetro *quantile*, el cual determina qué tan estrecha o amplia será la ventana de búsqueda de los clusters. Se utilizó un quantile = 0.02 como parámetro de entrada para estimar el ancho de banda de la ventana de densidad. Este valor fue elegido tras observar que valores mayores agrupaban demasiado los datos y ocultaban zonas de riesgo localizadas, mientras que valores

menores producían una segmentación excesiva e interpretable solo a microescala. El valor 0.02 ofreció un equilibrio adecuado entre sensibilidad a concentraciones delictivas pequeñas y coherencia visual de los resultados sobre el mapa.

La variable quantile puede ajustarse para controlar la sensibilidad del algoritmo en la detección de clusters. A diferencia de otras técnicas de clustering que requieren especificar el número de grupos, Mean Shift encuentra el número óptimo de clusters en función de la densidad de los datos, agrupando puntos cercanos y detectando los centros automáticamente. Esto permite una mayor flexibilidad, especialmente cuando no se tiene un conocimiento previo sobre la cantidad de clusters presentes en el conjunto de datos.

Clustering Jerárquico

El clustering jerárquico (hierarchical clustering) es un algoritmo de agrupamiento que organiza los datos en una estructura jerárquica basada en la distancia o similitud entre ellos. Su objetivo es maximizar la homogeneidad dentro de cada cluster. Este algoritmo presenta dos enfoques principales:

1. Aglomerativo (Bottom-Up): Cada dato comienza como un cluster independiente, y estos se van fusionando progresivamente en clusters más grandes según su similitud.
2. Divisivo (Top-Down): Se inicia con un único cluster que contiene todos los datos, y este se divide gradualmente en clusters más pequeños.

Vol. 10 No. 30 Septiembre-diciembre (2025)
ISSN: 2448-5497

Ambos enfoques resultan útiles para identificar patrones y relaciones jerárquicas en los datos, facilitando su interpretación y análisis(Castillo, 2023).

Este tipo de agrupamiento es especialmente útil para explorar relaciones jerárquicas y estructurales en los datos, representadas frecuentemente mediante diagramas llamados dendrogramas (Vichi, Cavicchia y Groenen, 2022).

El algoritmo de Clustering Jerárquico, utiliza varios parámetros clave para agrupar datos en función de su similitud. En este estudio, se empleó el método aglomerativo (*Agglomerative Clustering*) con el criterio de enlace de *Ward*, que minimiza la varianza total dentro de los grupos al combinar clusters. En este caso, se utilizó el enfoque aglomerativo con el método de enlace de Ward (linkage='ward') y se fijó n_clusters = 8 para mantener la comparación directa con K-means y DBSCAN. El método de Ward fue seleccionado por su capacidad para formar clusters compactos y balanceados, lo cual se ajustaba a la distribución de los datos observados. Además, el análisis en el dendrograma permitió validar que ocho agrupaciones generaban una segmentación razonable sin pérdida importante de estructura jerárquica.

Se evaluó el comportamiento de cada algoritmo considerando criterios como la cohesión de los clusters, la capacidad de detección de zonas de alta densidad delictiva y la facilidad de interpretación de los resultados. También se consideró la sensibilidad de cada algoritmo a los parámetros y su eficiencia computacional.

Se generaron mapas temáticos que muestran la distribución de los clusters obtenidos por cada algoritmo. Estas visualizaciones permitieron identificar zonas de alto riesgo y patrones delictivos que no eran evidentes en una revisión superficial.

Se compararon los resultados obtenidos por los distintos algoritmos, discutiendo sus ventajas y limitaciones en el contexto de datos espaciales de seguridad pública. Se propusieron interpretaciones de los patrones identificados y su posible utilidad para la toma de decisiones en seguridad ciudadana.

Se integraron todos los hallazgos, gráficas, mapas y análisis en un documento formal de investigación, redactando el artículo académico con base en las actividades descritas y los resultados obtenidos.

Para observar el agrupamiento de los datos, se utilizó un equipo con procesador AMD Ryzen 7 5700U y 16 GB de RAM con el entorno de desarrollo Spyder (anaconda3) con el lenguaje de programación Python en su versión 3.9 junto con las librerías numpy, matplotlib y Scikit-learn para el clustering. Para la representación de los resultados, se utilizaron herramientas de visualización en Python, particularmente la biblioteca Folium, que permitió construir mapas interactivos. Se superpusieron los clusters generados sobre un mapa de la Ciudad de México, incorporando además una capa de división política por alcaldías, obtenida mediante un archivo GeoJSON oficial.

Con el fin de mejorar la legibilidad del mapa, se incorporaron:

- Una escala de coordenadas latitud y longitud con ticks y etiquetas, colocada en la esquina inferior izquierda.
- Líneas guía en forma de escuadra para facilitar la orientación espacial del usuario.
- Colores diferenciados para cada cluster y para los puntos de ruido (no agrupados).
- Etiquetas emergentes con información del cluster asignado.

Vol. 10 No. 30 Septiembre-diciembre (2025)
ISSN: 2448-5497

Se incorporó una evaluación cuantitativa con el fin de comparar objetivamente el rendimiento de cada algoritmo. Para ello, se calcularon dos métricas estándar de validación interna:

- Coeficiente de Silhouette, el cual mide la cohesión y separación de los grupos generados. Esta métrica se basa en la comparación de la distancia media entre cada punto y los puntos de su mismo cluster con la distancia media entre dicho punto y los puntos del cluster más cercano. Su valor oscila entre -1 y 1, donde valores cercanos a 1 indican una mejor estructuración del agrupamiento (Liu, 2024).
- Índice de Davies-Bouldin (DBI), el cual evalúa la relación entre la dispersión intra cluster y la separación entre clusters. Un valor menor indica una mejor definición de los grupos, es decir, clusters más compactos y bien diferenciados entre sí (Fermín, 2024).

Estas métricas fueron aplicadas a los resultados obtenidos por cada algoritmo utilizando las etiquetas generadas en el proceso de agrupamiento. En el caso de DBSCAN, se identificaron múltiples puntos clasificados como ruido (etiquetados como -1). Para evaluar cuantitativamente la calidad de los clusters, se excluyeron estos puntos y se calcularon las métricas de validación interna solo con los datos asignados a clusters válidos. Este enfoque es adecuado, dado que las métricas como el coeficiente de Silhouette y el índice de Davies-Bouldin requieren al menos dos grupos bien definidos para proporcionar resultados significativos. La implementación se realizó con la biblioteca scikit-learn, específicamente mediante las funciones silhouette_score y davies_bouldin_score.

Resultados

Se aplicaron los cuatro algoritmos de clustering a los 862 registros de delitos tipo "robo a transeúnte en vía pública" ocurridos en diciembre de 2022 en la Ciudad de México. A continuación, se presentan los resultados obtenidos por cada algoritmo.

DBSCAN

DBSCAN identificó un total de 8 clusters, excluyendo varios puntos clasificados como ruido por su baja densidad local. En la Figura 1 se observa la capacidad de este algoritmo para generar agrupamientos de forma arbitraria, lo que permitió detectar zonas de alta concentración delictiva sin forzar una geometría regular. Los clusters se concentraron especialmente en las alcaldías de Cuauhtémoc, Iztapalapa y Venustiano Carranza, coincidiendo con zonas urbanas de alta densidad y movilidad. Alcaldías donde hay mayor presencia de puntos.

- Cluster 1 (rojo): Azcapotzalco, Gustavo A madero, Cuauhtémoc y Venustiano Carranza.
- Cluster 2 (azul): Álvaro Obregón y Benito Juárez.
- Cluster 3 (verde): Benito Juárez.
- Cluster 4 (negro): Venustiano Carranza.

Vol. 10 No. 30 Septiembre-diciembre (2025)
ISSN: 2448-5497

- Cluster 5 (amarillo): Iztapalapa.
- Cluster 6 (azul oscuro): Iztapalapa.
- Cluster 7 (marrón): Coyoacán.
- Cluster 8 (violeta): Coyoacán.

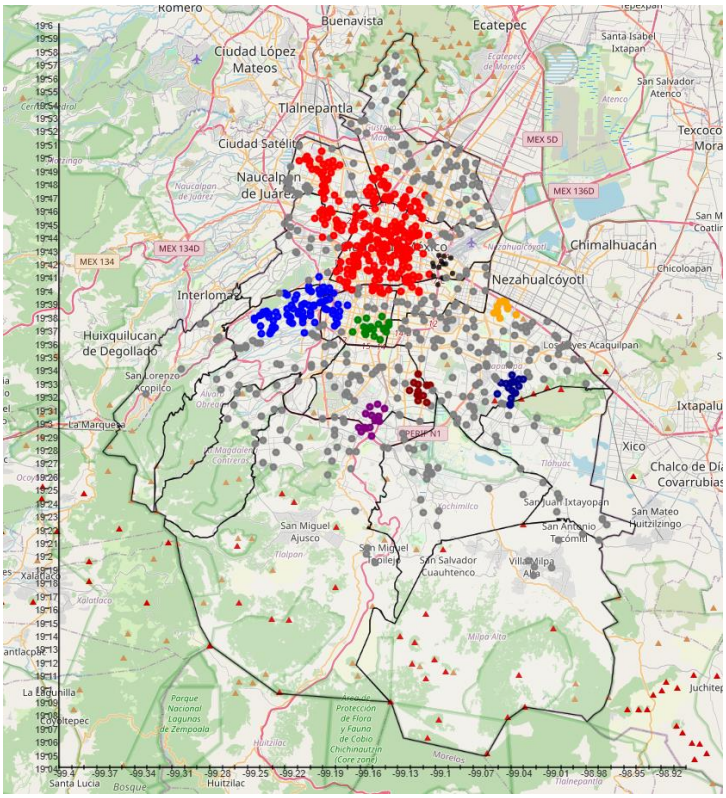


Figura 1. Aplicación del algoritmo DBSCAN para delitos registrados en Ciudad de México.

K-means

El algoritmo agrupó los datos en 8 clusters predefinidos, con formas más homogéneas, pero menos precisas en los bordes. Las zonas resultantes fueron más simétricas, pero menos sensibles a variaciones de densidad. Se observó concentración de delitos en áreas similares a

DBSCAN, aunque con mayor dispersión. El resultado del agrupamiento se observa en la Figura 2. Las alcaldías de donde hay mayor presencia de puntos se encuentran en los siguientes agrupamientos:

- Cluster 1 (amarillo): Gustavo A Madero.
- Cluster 2 (azul oscuro): Azcapotzalco y Miguel Hidalgo.
- Cluster 3 (azul): Cuauhtémoc, Iztacalco, Venustiano Carranza.
- Cluster 4 (violeta): Álvaro Obregón, Cuauhtémoc, Benito Juárez.
- Cluster 5 (marrón): Iztapalapa.
- Cluster 6 (verde): Coyoacán, Tlalpan, Xochimilco.
- Cluster 7 (negro): Cuajimalpa de Morelos, Álvaro Obregón, Magdalena de contreras y Tlalpan.
- Cluster 8 (rojo): Con pocos puntos distribuidos entre Tláhuac, Xochimilco, Milpa Alta.

DIOTIMA
REVISTA CIENTÍFICA DE ESTUDIOS TRANSDISCIPLINARIA
Vol. 10 No. 30 Septiembre-diciembre (2025)
ISSN: 2448-5497

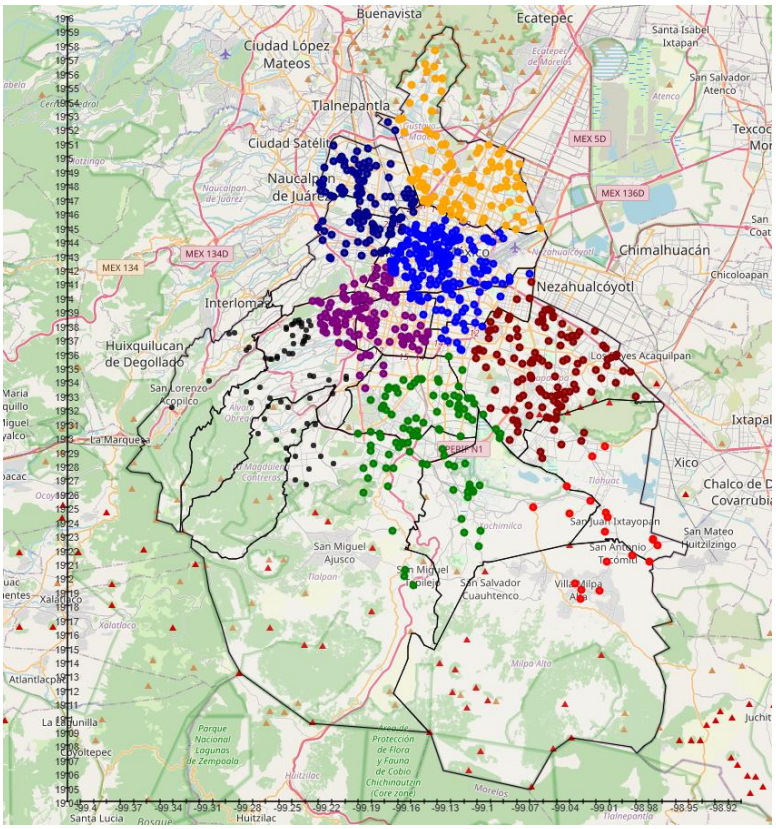


Figura 2. Aplicación del algoritmo K-means para delitos registrados en Ciudad de México.

Mean Shift

El algoritmo detectó 48 clusters automáticamente sin definir previamente el número de grupos. Se observaron agrupamientos más finos, lo que permitió detectar zonas pequeñas con alta incidencia delictiva, pero con demasiados puntos dispersos considerados como ruido (ver Figura 3). Fue el algoritmo más sensible a variaciones de densidad, aunque más costoso computacionalmente. Los cluster con mayor densidad se encuentran en las alcaldías Azcapotzalco, Gustavo A Madero, Venustiano Carranza, Miguel Hidalgo, Álvaro Obregón, Benito Juárez, Iztacalco, Iztapalapa, Coyoacán, Tlalpan, Xochimilco y Cuauhtémoc.

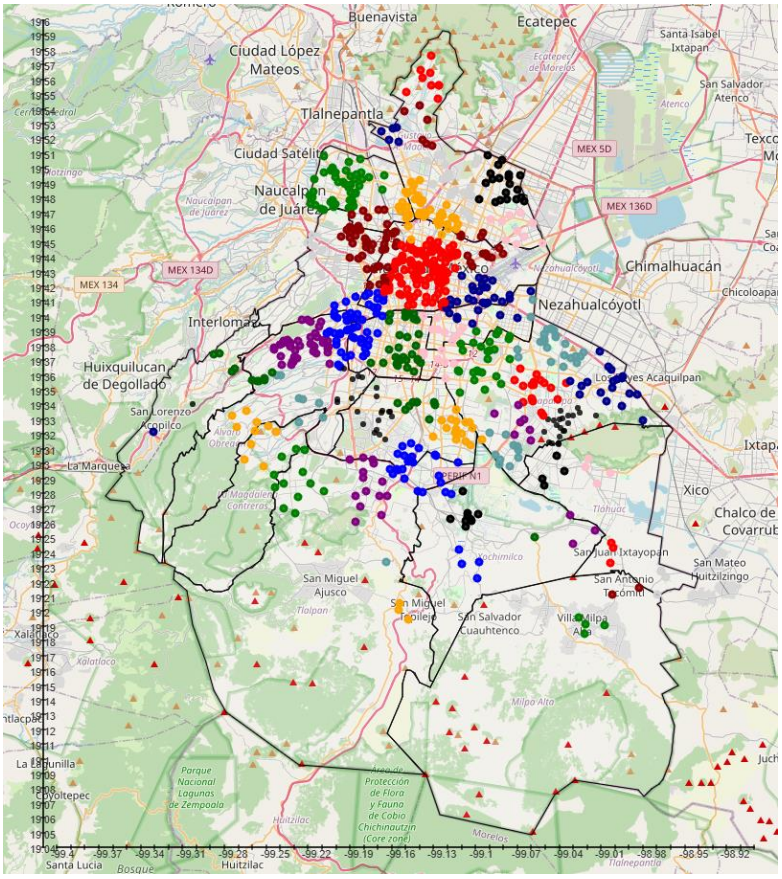


Figura 3. Aplicación del algoritmo Mean shift para delitos registrados en Ciudad de México.

Clustering jerárquico

Se generaron 8 clusters utilizando el método de Ward, con resultados similares en forma a K-means, pero con mayor separación entre grupos (ver Figura 4). En la Figura 5 se puede observar el dendrograma con la relación jerárquica entre grupos, mostrando que algunos clusters estaban más estrechamente relacionados. La distribución de los cluster en las diferentes alcaldías es la siguiente:

Vol. 10 No. 30 Septiembre-diciembre (2025)
ISSN: 2448-5497

- Cluster 1 (verde): Gustavo A Madero y Venustiano Carranza.
- Cluster 2 (marrón): Azcapotzalco y Miguel Hidalgo.
- Cluster 3 (rojo): Cuauhtémoc, Venustiano Carranza, Iztacalco e Iztapalapa.
- Cluster 4 (amarillo): Miguel Hidalgo, Cuauhtémoc, Álvaro Obregón, Benito Juárez y Coyoacán.
- Cluster 5 (violeta): Iztapalapa y Tláhuac.
- Cluster 6 (azul): Coyoacán, Tlalpan y Xochimilco.
- Cluster 7 (negro): Álvaro Obregón, y con pocos puntos distribuidos entre Cuajimalpa de Morelos, La Magdalena Contreras y Tlalpan.
- Cluster 8 (azul oscuro): Con pocos puntos distribuidos entre Xochimilco Tláhuac y Milpa Alta.

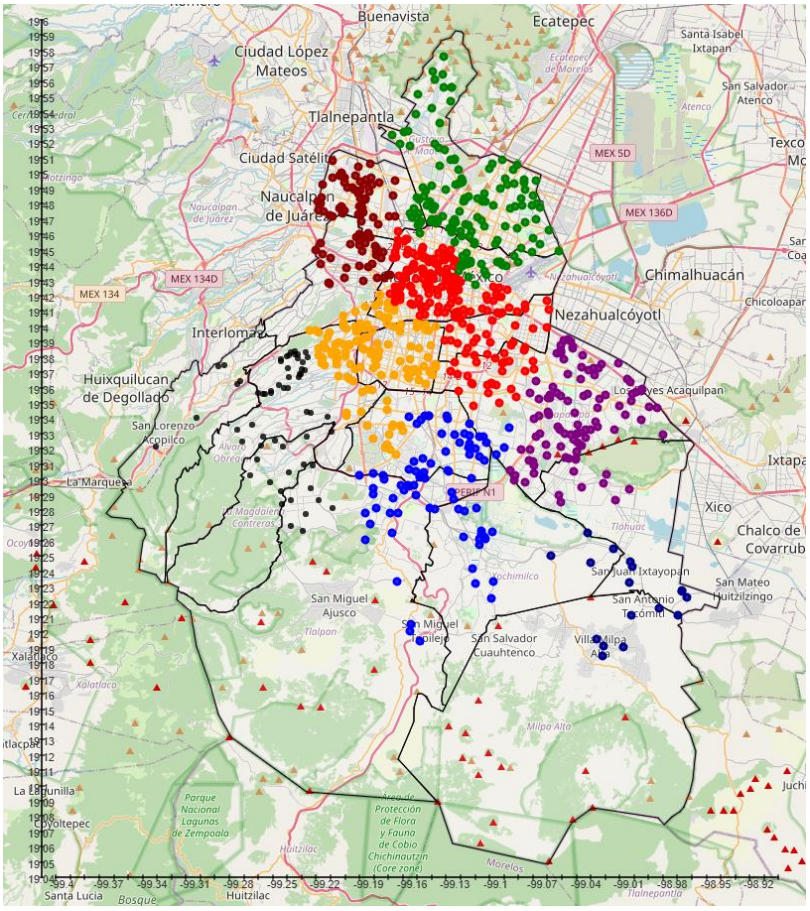
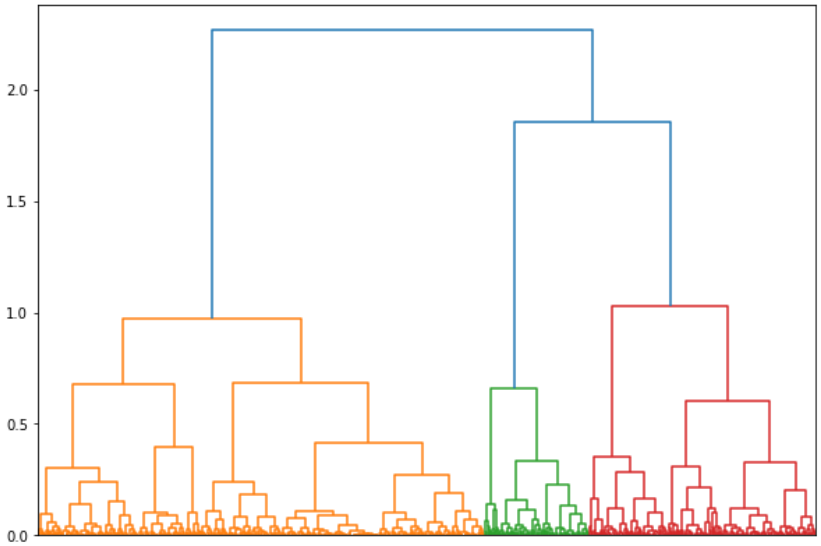


Figura 4. Aplicación del algoritmo clustering jerárquico aglomerativo para delitos registrados en Ciudad de México.



Vol. 10 No. 30 Septiembre-diciembre (2025)
ISSN: 2448-5497

Figura 5. Dendrograma del algoritmo clustering jerárquico aglomerativo para delitos registrados en Ciudad de México.

Discusión

La aplicación de algoritmos de clustering en datos espaciales, como DBSCAN, K-means, Mean Shift y el Clustering Jerárquico, revela diferencias significativas en su comportamiento y resultados. Mientras que DBSCAN no agrupa todos los elementos debido a la dispersión de los datos y los parámetros establecidos, generando automáticamente un menor número de grupos y detectando valores atípicos con mayor facilidad (Wororomi, Allo y Paranoan, 2023), K-means requiere especificar el número de clusters y tiende a incluir todos los elementos en el proceso de agrupación, lo que puede resultar en agrupaciones más compactas y coherentes según ciertos índices de validación interna como el coeficiente de Silhouette (Tomašev y Radovanović, 2015; Wororomi et al., 2023). Por su parte, el Clustering Jerárquico permite visualizar la estructura jerárquica de los datos, proporcionando una representación más detallada de las relaciones entre clusters, aunque también requiere definir el número de grupos (Palacio-Niño y Berzal, 2019). En contraste, Mean Shift detecta automáticamente el número de clusters basándose en la densidad de los datos, lo que lo hace flexible para identificar agrupaciones sin requerir parámetros como el número de clusters. Estas diferencias indican cómo la selección del algoritmo y la interpretación de sus resultados pueden depender en gran medida de la naturaleza de los datos y de las métricas de evaluación utilizadas, especialmente en contextos geoespaciales donde la densidad y la distribución de los puntos pueden influir notablemente en los agrupamientos obtenidos (Tomašev y Radovanović, 2015; Palacio-Niño y Berzal, 2019).

Para facilitar la comparación entre los algoritmos de clustering evaluados, se presenta en la Tabla 1 un resumen que incluye el número de clusters detectados, la capacidad de manejo de ruido, la forma característica de los clusters generados, así como el coste computacional aproximado observado durante la ejecución de los algoritmos.

Algoritmo	**Nº de Clusters**	**Detección de Ruido**	**Forma de los Clusters**	**Comentarios principales**	**Tiempo de ejecución**
DBSCAN	8	Sí	Irregulares	Alta precisión en zonas densas; sensible a eps	0.3 segundos
K-means	8 (predefinido)	No	Simétricas	Rápido y eficiente; requiere definir número de clusters	0.1 segundos
Mean Shift	12	No	Ajustadas por densidad	No requiere número de clusters; detecta zonas pequeñas, alto coste	1.2 segundos
Jerárquico	8	No	Compactas	Analiza relaciones entre grupos; útil visualmente (dendrograma)	0.7 segundos

Tabla 1. Comparación entre algoritmos de clustering aplicados.

Vol. 10 No. 30 Septiembre-diciembre (2025)
ISSN: 2448-5497

En cuanto a la evaluación cuantitativa como se observa en la Tabla 2, el algoritmo K-means obtuvo el mayor coeficiente de Silhouette (0.399), lo cual sugiere que los puntos dentro de cada cluster están bien cohesionados y suficientemente separados de otros clusters. Sin embargo, también presentó un índice de Davies-Bouldin relativamente alto (0.781), indicando que algunos clusters podrían estar más próximos entre sí. DBSCAN registró un coeficiente de Silhouette de 0.338, menor que K-means, pero alcanzó el mejor índice Davies-Bouldin (0.466), lo que refleja una menor superposición entre los grupos y una mayor compacidad relativa. Mean Shift y el clustering jerárquico (Ward) presentaron resultados intermedios en ambas métricas, con coeficientes de Silhouette de 0.362 y 0.359 respectivamente, y valores de Davies-Bouldin de 0.691 y 0.819, lo que indica un desempeño moderado en cuanto a separación y coherencia interna. En general, los resultados muestran que DBSCAN destaca por la separación entre clusters (aunque hay que considerar que los cluster son mas pequeños y existen demasiados puntos con ruido), mientras que K-means ofrece la mejor cohesión interna, aunque a costa de cierta cercanía entre grupos.

Algoritmo	**Coeficiente Silhouette**	**Índice Davies-Bouldin**
DBSCAN	0.3377055406238473	0.46589080358552376
K-means	0.3991982835695953	0.780557053747801
Mean Shift	0.3622010360867173	0.69182274174502
Jerárquico	0.3594366915785989	0.8195417025032317

Tabla 2. Métricas de evaluación cuantitativa para cada algoritmo de clustering.

Conclusiones

Los hallazgos de este estudio abren la posibilidad de desarrollos que mejoren la identificación de patrones de delincuencia mediante algoritmos de clustering, adaptando los procesos que utilizan a las características específicas de los datos. Es importante mencionar las limitaciones, como la sensibilidad de los resultados a los parámetros seleccionados y la calidad de los datos. El uso de algoritmos de clustering para la clasificación de información espacial representa un avance significativo en el análisis de datos relacionados con la delincuencia.

Como propuesta final, se plantea el desarrollo de un algoritmo híbrido de clustering, concebido como un meta-algoritmo adaptativo capaz de integrar las fortalezas de los enfoques analizados (ver Figura 6). La idea central es que este sistema evalúe las características estadísticas y espaciales del conjunto de datos (como densidad, dispersión o ruido), y con base en estos criterios, seleccione el algoritmo de clustering más adecuado (por ejemplo, DBSCAN para datos densos y con ruido, o K-means para estructuras más uniformes). Es posible combinar resultados parciales de varios algoritmos, generando una clasificación final más robusta. Este algoritmo se conceptualiza para su integración en sistemas de información geográfica (SIG), facilitando la segmentación de territorios urbanos de acuerdo con el nivel de riesgo detectado, con aplicaciones directas en la planificación de estrategias de seguridad pública.

Vol. 10 No. 30 Septiembre-diciembre (2025)
ISSN: 2448-5497

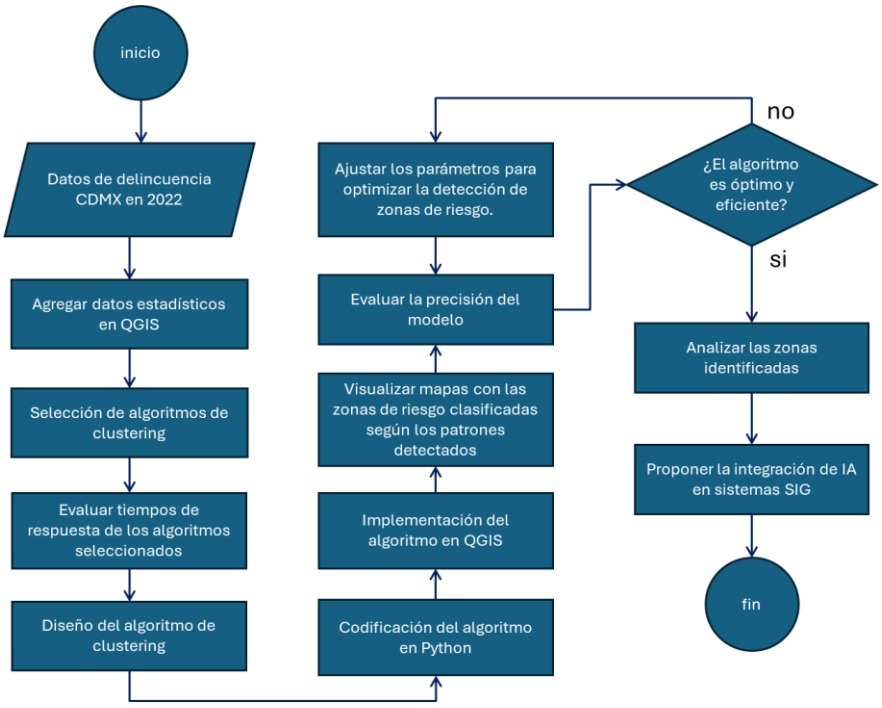


Figura 6. Diseño la elaboración de algoritmo propuesto.

En función de las métricas de validación interna aplicadas (coeficiente de Silhouette e índice de Davies-Bouldin), el algoritmo K-means obtuvo los mejores resultados comparativos, en el coeficiente de Silhouette y en el índice de Davies-Bouldin, lo que sugiere una buena separación entre clusters y una aceptable compacidad interna. Aunque DBSCAN mostró una mejor puntuación en el índice de Davies-Bouldin, su bajo valor de Silhouette refleja menor cohesión, posiblemente influido por la presencia de puntos clasificados como ruido. Mean Shift y Clustering Jerárquico presentaron resultados intermedios. Esta evaluación permite concluir que, aunque cada algoritmo tiene fortalezas particulares (como la capacidad de DBSCAN para excluir ruido o la adaptabilidad de Mean Shift), K-means ofrece el mejor rendimiento general bajo los criterios de validación empleados. Esta evidencia justifica su inclusión como componente clave dentro del diseño propuesto del

algoritmo híbrido, en combinación con otros métodos que complementen sus limitaciones estructurales.

Vol. 10 No. 30 Septiembre-diciembre (2025)
ISSN: 2448-5497

Referencias

Amat, J. (2017). Clustering y heatmaps: aprendizaje no supervisado. Recuperado el 15 de octubre de 2023 de https://goo.su/RvZGkMS

Carrasco, I., J. y Moyotl, H., E. (2022). Agrupamiento basado en densidad para la detección automática de hot spots delictivos en la CDMX. *Research in Computing Science*, 151(8). https://doi.org/10.47741/17943108.123

Castillo, G. (2023). Aprendizaje no supervisado: ¿Qué es y cómo funciona?. Recuperado el 21 de septiembre de 2023 de https://goo.su/BQzS

Barreno R., P., Bregón B., A., y Martínez P., M. (2021). *Estudio de técnicas de clustering y detección de anomalías aplicado a fresadoras CNC* [Trabajo de grado, Universidad de Valladolid]. Escuela de Ingeniería Informática de Segovia.

Fermín G., R. F. (2024). *Revisión de métodos de clustering para datos funcionales* [Trabajo de fin de máster, Universidad de A Coruña]. Máster en Técnicas Estadísticas. http://eio.usc.es/pub/mte/descargas/ProyectosFinMaster/Proyecto_2176.pdf

Fontalvo, H., T. J., Vega, H., M. A., y 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

González, L. (2020). Algoritmos de Agrupamiento. AprendeIA. Recuperado el 14 de mayo de 2024 de https://aprendeia.com/algoritmos-de-clustering-agrupamiento-aprendizaje-no-supervisado/

Liu, G. (2024). *A new index for clustering evaluation based on density estimation*. arXiv. https://arxiv.org/abs/2207.01294

Mohammad, N. (2023). A Computational Theory and Semi-Supervised Algorithm for Clustering. *arXiv, cs.LG*. https://doi.org/10.48550/arXiv.2306.06974

Palacio-Niño, J.-O., y Berzal, F. (2019). *Evaluation metrics for unsupervised learning algorithms*. arXiv. https://arxiv.org/abs/1905.05667

Pedrero, V., Reynaldos-Grandón, K., Ureta-Achurra, J., y Cortez-Pinto, E. (2021). Generalidades del Machine Learning y su aplicación en la gestión sanitaria en Servicios de Urgencia. *Revista Médica de Chile*, 149(2), 248-254. http://dx.doi.org/10.4067/s0034-98872021000200248

Ramírez, L. (2023). Algoritmo k-means: ¿Qué es y cómo funciona? IEBS. Recuperado el 13 de mayo de 2024 de https://www.iebschool.com/blog/algoritmo-k-means-que-es-y-como-funciona-big-data/

Tomašev, N., & Radovanović, M. (2016). Clustering evaluation in high-dimensional data. In J. A. Lee & M. Verleysen (Eds.), *High-dimensional data analysis with low-dimensional models* (pp. 71–107). Springer. https://doi.org/10.1007/978-3-319-24211-8_4

Unidad de Estadística y Transparencia de la Ciudad de México. (2025). *Boletín estadístico de la incidencia delictiva en la Ciudad de México del mes de marzo 2025*. Fiscalía General de Justicia de la Ciudad de México. Recuperado el 23 de abril de 2025 de https://www.fgjcdmx.gob.mx/storage/app/media/Estadisticas%20Delictivas/2025/03-boletin-marzo-2025.pdf

Vol. 10 No. 30 Septiembre-diciembre (2025)
ISSN: 2448-5497

Vera, G. (2021). *Desarrollo de un algoritmo evolutivo para la perfilación geográfica criminal*. Tesis de doctorado. Universidad Autónoma del Estado de México

Vichi, M., Cavicchia, C., y Groenen, P. J. F. (2022). Hierarchical Means Clustering. *Journal of Classification*, 39, 553–577. https://doi.org/10.1007/s00357-022-09419-7

Wororomi, J., Allo, C., y Paranoan, N. (2023). Performance of K-Means and DBSCAN algorithm in clustering gross regional domestic product. *Journal of International Conference Proceedings*, 6(5), 179–193. https://doi.org/10.32535/jicp.v6i5.2710