



Evaluación comparativa del clustering jerárquico y k-means para la segmentación estratégica de clientes en empresas comerciales.

Comparative evaluation of hierarchical clustering and k-means for strategic customer segmentation in commercial companies.

Carlos Bladimir Moreano Guerra*

moreanocarlos1@gmail.com

Tania Eslavenska Escobar Erazo*

taniaemy8@yahoo.es

Luis Fernando Herrera Moreno**

ferherse@gmail.com

Julio Andrés Escobar Cardenas***

jaescobar6018@gmail.com

* Universidad Central del Ecuador, **Universidad Politécnica Salesiana, ***Universidad de las Fuerzas Armadas, Quito, Ecuador

Correspondencia: moreanocarlos1@gmail.com

RESUMEN

En el ámbito empresarial actual, la segmentación de clientes es un recurso clave para optimizar tanto la toma de decisiones como el diseño de estrategias comerciales sustentadas en el análisis de información. En este contexto, las técnicas de clustering han adquirido relevancia, debido a que permiten identificar patrones de comportamiento, además de agrupar clientes en base a sus hábitos de consumo. Este estudio tiene como objetivo evaluar comparativamente el rendimiento del clustering jerárquico y K-means para la segmentación estratégica de clientes en empresas del sector comercial, considerando como variables la frecuencia de compras, el monto de las transacciones, así como la antigüedad del cliente. En la investigación se aplicó un enfoque cuantitativo basado en la analítica de datos; iniciando con el preprocesamiento y normalización de las variables. Posteriormente, se determinó el número óptimo de clústeres mediante el método del codo, se empleó el clustering jerárquico para el diagnóstico exploratorio de la estructura de los datos, seguido del algoritmo k-means para efectuar la clasificación final de grupos homogéneos. Adicionalmente se recurrió al índice silhouette como métrica de validación interna para examinar la calidad de los conjuntos obtenidos. Los resultados revelaron la existencia de segmentos de clientes con patrones conductuales diferenciados, logrando distinguir perfiles de consumo frecuentes y ocasionales. En conclusión, la implementación combinada de estas técnicas posibilita la obtención de segmentaciones confiables y coherentes que sirven de apoyo para la toma de decisiones estratégicas, así como para el desarrollo de tácticas de mercado personalizadas.

Palabras claves: Segmentación de Clientes, K-means, Clustering Jerárquico, Comportamiento del Cliente, Análisis de Clúster.

ABSTRACT

In today's business environment, customer segmentation is a key resource for optimizing both decision-making and the design of business strategies based on data analysis. In this context, clustering techniques have gained relevance because they allow for the identification of behavioral patterns and the grouping of customers based on their consumption habits. This study aims to comparatively evaluate the performance of hierarchical clustering and K-means for the strategic segmentation of customers in companies in the retail sector, considering purchase frequency, transaction value, and customer tenure as variables. The research employed a quantitative approach based on data analytics, beginning with the preprocessing and normalization of the variables. Subsequently, the optimal number of clusters was determined using the elbow method. Hierarchical clustering was used for the exploratory diagnosis of the data structure, followed by the K-means algorithm for the final classification of homogeneous groups. Additionally, the Silhouette index was used as an internal validation metric to examine the quality of the resulting clusters. The results revealed the existence of customer segments with distinct behavioral patterns, distinguishing between frequent and occasional consumers. In conclusion, the combined implementation of these techniques enables the creation of reliable and consistent segmentations that support strategic decision-making and the development of personalized marketing tactics.

Keywords: Customer Segmentation, K-means, Hierarchical Clustering, Customer Behavior, Cluster Analysis.

Cómo citar:

Moreano Guerra, C. B., Escobar Erazo, T. E., Herrera Moreno, L. F., & Escobar Cardenas, J. A. (2026). Evaluación comparativa del clustering jerárquico y k-means para la segmentación estratégica de clientes en empresas comerciales. GADE: Revista Científica, 6(1), 674-703. <https://doi.org/10.63549/rg.v6i1.806>



INTRODUCCIÓN

En el entorno empresarial actual, caracterizado por su alta competitividad, las empresas comerciales demandan instrumentos de analítica que les permita descifrar el comportamiento de sus clientes y perfeccionar sus estrategias de negocio. En este contexto, la segmentación de clientes representa un pilar estratégico, dado que segmentar el mercado es una respuesta a la variedad de preferencias e intereses de los consumidores. Las empresas, al no poder cubrir todas las demandas, optan por clasificar su clientela en segmentos específicos con la finalidad de incrementar su eficiencia operativa y potenciar su rentabilidad (Camargo Morales et al., 2023)

Bajo el paradigma de la economía digital contemporánea, las empresas se enfrentan al imperativo desafío de proveer experiencias personalizadas en cada punto de contacto con los clientes, quienes exigen relaciones fluidas, relevantes y adaptadas a sus preferencias, lo cual ha restado efectividad a las tácticas de marketing genéricas.

En este sentido, la gestión de la interacción con los clientes se ha erigido como una prioridad corporativa crítica,

debido a que contribuye a optimizar la satisfacción del consumidor, reforzar la lealtad y maximizar la rentabilidad empresarial (Alves Gomes & Meisen, 2023).

En el contexto competitivo actual, la detección de segmentos de clientes representa un factor determinante tanto para optimizar la calidad del servicio, como para reforzar la satisfacción de los clientes, además de promover relaciones comerciales sustentables. Bajo este enfoque, la segmentación de clientes se ha afianzado como una herramienta estratégica que posibilita entender a profundidad las motivaciones e intereses de compra de los consumidores, promoviendo el diseño de estrategias comerciales más efectivas y dirigidas hacia la fidelización.

De acuerdo con Prieto Herrera (2021) en el ámbito empresarial la segmentación de clientes conlleva la clasificación de compradores en conjuntos homogéneos con comportamientos y características similares; este proceso resulta esencial tanto para la personalización de las estrategias de marketing, como para optimizar la satisfacción de los clientes e incrementar las ventas.



Bajo este enfoque, la segmentación de clientes, no solo facilita la gestión de específicos, sino que además constituye la base analítica necesaria para la implementación de técnicas cuantitativas orientadas a la identificación de patrones de comportamiento. Las organizaciones en la actualidad generan un gran volumen de datos, lo cual ha impulsado la aplicación de técnicas estadísticas como minería de datos y aprendizaje automático como herramienta en los procesos de análisis estratégico.

Debido a que mediante el data mining o minería de datos es posible analizar aquellos factores que afectan diversos procesos, anticipar escenarios futuros, además de segmentar elementos con características similares. Su principal ventaja consiste en detectar relaciones modelos y patrones implícitos en los datos sin depender de reglas predefinidas, lo cual potencia la capacidad predictiva de la empresa y favorece una gestión proactiva basada en la evidencia de datos (Uvidia Fassler et al., 2018).

Por su parte el aprendizaje automático como rama de la inteligencia artificial, está orientado a la programación de modelos capaces de

identificar vínculos y regularidades en los conjuntos de datos; esta capacidad posibilita inferir futuros comportamientos, además de reforzar la gobernanza de datos en las empresas (Navarrete Vincés et al., 2025).

El empleo de técnicas analíticas y herramientas estadísticas ha reforzado los procesos de segmentación de clientes, posibilitando la obtención de resultados más objetivos, precisos, así como con rigor científico. De este modo, la integración de avanzados métodos analíticos coadyuva a la eficacia estratégica de las empresas, facilitando una detallada interpretación tanto de la arquitectura como de la evolución de los segmentos, lo cual fortalece la toma de decisiones fundamentada en datos.

Dado al exponencial crecimiento de datos, las empresas requieren herramientas de máxima eficiencia para procesar información a gran escala. Mediante la implementación de algoritmos especializados, resulta posible detectar tanto comportamientos como tendencias claves. El aprendizaje automático es el principal instrumento para agrupar datos y categorizar clientes facilitando acciones estratégicas fundamentadas en el análisis real de los datos (Moharana & Sarmah, 2018).



Bajo este enfoque, las técnicas de aprendizaje automático se clasifican en función tanto de la problemática como de las características de los datos. Entre las que destacan el aprendizaje supervisado y no supervisado; los cuales se caracterizan principalmente porque el primero emplea datos con etiquetas para predecir resultados específicos, mientras que el segundo analiza información sin etiquetar y sin categorizar, para identificar estructuras o patrones ocultos de forma autónoma, por esta razón es utilizado comúnmente en la segmentación de datos, reducción de dimensionalidad y agrupamiento (Stack Sánchez, 2021).

Para la segmentación de clientes, el aprendizaje no supervisado resulta especialmente relevante, debido a que permite identificar segmentos intrínsecos en los datos, a partir de indicadores de comportamiento como la frecuencia de compras, monto de la transacción y la antigüedad del cliente. Este enfoque al no utilizar etiquetas, posibilita detectar patrones inéditos que sirven de soporte para la toma de decisiones estratégicas en las empresas.

Entre los métodos de aprendizaje no supervisado, una técnica que ha demostrado ser una herramienta eficaz

para la detección de tendencias implícitas en los datos prescindiendo de etiquetas o clasificaciones previas es el algoritmo de clustering, el cual es una herramienta analítica que agrupa elementos en función de sus similitudes, con el objetivo de lograr una homogeneidad interna alta y una marcada heterogeneidad externa, ayudando además a descifrar arquitecturas de datos difíciles de interpretar, consolidándose como un recurso estándar en múltiples sectores estratégicos (Franco-Árcega et al., 2021).

Dentro del aprendizaje no supervisado, el clustering se posiciona como una herramienta clave por su capacidad para detectar estructuras ocultas en los datos, sin requerir información previa sobre las categorías existentes. Esta técnica es ampliamente utilizada debido a que permite agrupar en un mismo conjunto aquellas observaciones con características similares, facilitando de este modo la identificación de patrones de comportamiento.

Conforme a lo manifestado por Alonso et al. (2024) en el entorno empresarial, especialmente en el área de marketing, el clustering es clave, debido



a que permite agrupar clientes con similares características, para lograr una eficiente gestión de la información. Esta técnica facilita la comprensión del comportamiento de la clientela a través de la segmentación, lo cual posibilita diseñar efectivas estrategias evitando el análisis individualizado de cada cliente.

La aplicación del clustering en la segmentación resulta muy útil, dado que posibilita no solo organizar, sino también simplificar enormes volúmenes de datos, además de optimizar la interpretación de la información y respaldar la toma de decisiones fundamentada en evidencias. En consecuencia, su implementación es transversal en la analítica empresarial, donde la formación de grupos homogéneos es crítica para la efectividad de las estrategias.

En el marco del aprendizaje no supervisado, coexisten diversos algoritmos de clúster diseñados para agrupar datos basándose en sus características y niveles de similitud. Estos métodos se clasifican en función de su lógica de agrupación, destacando particularmente los enfoques jerárquicos y no jerárquicos.

En el análisis de clustering, los métodos jerárquicos se categorizan en

estrategias ascendentes o aglomerativas y descendentes o divisivas, las cuales estructuran la información mediante la progresiva fusión o división de las observaciones en función de su nivel de similitud. Estos procedimientos proporcionan una representación gráfica a través de un dendograma, el cual es una herramienta que facilita tanto la identificación de la estructura de agrupamiento, como la estimación del número de clústeres. A diferencia de estos, los métodos no jerárquicos, como los algoritmos de partición distribuyen los datos en una cantidad predefinida de clústeres, mediante la minimización de las varianzas intra-clúster y la maximización de la distancia entre grupos. Ambos enfoques son muy aplicados en investigaciones empíricas debido a la capacidad que tienen para detectar tanto estructuras como patrones subyacentes en entornos de datos complejos (Ayala-Aldana & Parra-Cid, 2023).

Entre los métodos más aplicados para la segmentación de clientes, se encuentran el clustering jerárquico y el algoritmo K-meanas. El análisis de clustering jerárquico o HCA (por sus siglas en inglés), se define como una metodología de agrupación que



organiza las observaciones mediante un esquema jerárquico basado en sus similitudes. Su relevancia en el campo del aprendizaje no supervisado reside en su capacidad para detectar estructuras intrínsecas en diferentes tipos de datos y contextos, favoreciendo tanto el estudio exploratorio como la decodificación de patrones complejos (Wani, 2024).

Dentro del estudio exploratorio, el clustering jerárquico se posiciona como un método valioso, debido a su capacidad para identificar progresivamente estructuras intrínsecas, prescindiendo de una delimitación inicial del número de segmentos.

Según lo expresado por Koprřivová & Matušřinská (2023) una de las ventajas principales del clustering jerárquico radica en que no exige conocer con anterioridad la cantidad de grupos, lo cual suele implicar complejos procesos de cálculo. Mediante el dendograma el investigador puede establecer el número óptimo de clústeres para posteriormente robustecer el estudio con técnicas complementarias.

Esta herramienta destaca por su capacidad de visualizar las relaciones de similitud en los datos, a través de representaciones gráficas como el dendograma, lo que optimiza la

comprensión de las estructuras detectadas. Debido a ello, es utilizado habitualmente como un diagnóstico preliminar, antes de implementar algoritmos de partición como K-means, aportando a la obtención de segmentaciones más consistentes.

K-means por su parte, constituye un algoritmo de aprendizaje no supervisado, diseñado para segmentar información en k conjuntos homogéneos en función de sus características. Este método busca potenciar la cohesión interna y resaltar la disparidad entre grupos, a través de la optimización de la distancia entre los segmentos y sus centroides. Debido tanto a su alto rendimiento algorítmico, como a su idoneidad para el tratamiento de datos cuantitativos, representa una solución estándar para el procesamiento de enormes flujos de información, operando principalmente sobre escalas numéricas (Guañuna Viteri, 2022).

El algoritmo K-means destaca como uno de los métodos de particionamiento más aplicado debido a su agilidad algorítmica, facilidad de implementación, así como también por proporcionar resultados comprensibles. Su capacidad para organizar grandes volúmenes de datos en conjuntos con



características similares la convierte en una herramienta de gran valor tanto para el análisis exploratorio como para la segmentación.

Dada a la eficiencia y simplicidad que caracteriza al K-means, es frecuentemente implementado en el entorno empresarial para estructurar bases de datos, segmentar clientes, optimizar la gestión de inventarios y analizar el mercado. Este enfoque permite a las organizaciones alcanzar una mayor eficacia en sus procesos, posibilitando el descubrimiento de patrones que apoyen una toma de decisiones informada ante un masivo conjunto de datos (Ufeli et al., 2025).

Es necesario resaltar que el desempeño del K-means depende de la correcta selección del número de segmentos, además de contar con datos de calidad. Dado que según Abdel-Hakim et al. (2024) el éxito de la segmentación está condicionado por la habilidad del sistema para diferenciar perfiles en un complejo entorno de datos. Este rendimiento habitualmente se mide mediante un criterio de optimización que el algoritmo ajusta con la finalidad de lograr una adecuada formación de los clústeres.

Por lo tanto, es recomendable emplear herramientas de comprobación adicionales con la finalidad de garantizar que los segmentos identificados sean confiables y coherentes.

Distintas investigaciones han ratificado la eficacia de las técnicas de clustering para la segmentación de clientes en el ámbito empresarial. Un estudio realizado por Tabianan et al. (2022) aplicó algoritmos de clustering para segmentar clientes en el entorno del comercio electrónico. Los resultados resaltaron la eficiencia de K-means en la identificación de conjuntos homogéneos, así como en la categorización de perfiles según su rentabilidad y comportamiento, corroborando la pertinencia de esta técnica como soporte no solo en la toma de decisiones, sino incluso en el diseño de estrategias comerciales sustentadas en datos.

Una investigación elaborada por Trujillo Logroño (2025) evidenció la eficacia del K-means como recurso para la segmentación de clientes, debido a que su implementación posibilita la comprensión tanto del valor, como de las necesidades de los usuarios, aportando a la creación de estrategias de marketing más personalizadas, al aprovechamiento



óptimo de los recursos y a una adecuada toma de decisiones empresariales.

Otros estudios como el desarrollado por Afzal et al. (2024) implementó el clustering jerárquico para la segmentación de clientes en el sector retail, con la finalidad de detectar patrones conductuales con base en variables sociodemográficas y hábitos de consumo. La validación mediante dendogramas, gráficos de dispersión e índice de silueta, corroboró la efectividad de esta técnica para revelar tanto estructuras de clústeres, como también relaciones entre los diversos segmentos de clientes. Los resultados no solo optimizaron la interpretación de las preferencias del mercado, sino que además fundamentaron el diseño de estrategias de marketing personalizadas, resaltando la relevancia del clustering jerárquico como herramienta para el análisis exploratorio de los datos, y para la toma de decisiones corporativas apoyada en datos.

Estudios recientes han analizado la aplicación conjunta del clustering jerárquico y el algoritmo K-means para la segmentación de clientes, como el efectuado por Singh et al. (2022) quienes evidencian en su investigación que la combinación de ambas técnicas permite

emplear la eficiencia del K-means para la creación de los grupos iniciales y la capacidad del clustering jerárquico tanto para profundizar como para comprender mejor la estructura de los datos. Los resultados obtenidos demostraron que este enfoque mixto no solo optimiza la interpretabilidad de los conjuntos, sino que además refuerza la toma de decisiones fundamentada en los datos analizados.

Las investigaciones revisadas ratifican la importancia de aplicar métodos de aprendizaje no supervisado, tales como el clustering jerárquico y el K-means, en el análisis de datos empresariales. Sin embargo, pese a su extendida implementación y las contribuciones de estas técnicas, persiste aún la necesidad de profundizar en estudios comparativos que examinen su desempeño en entornos de datos reales, además de que evalúen su eficacia como soporte en la segmentación estratégica de clientes en empresas del sector comercial.

Bajo este contexto, el presente estudio tiene como objetivo evaluar comparativamente la aplicación de los métodos de clustering jerárquico y K-mean para la segmentación estratégica de clientes en empresas comerciales, a



través del análisis de variables de comportamiento como la frecuencia de compras, el monto de las transacciones y la antigüedad de los compradores. Adicionalmente se evaluará tanto la consistencia como la confiabilidad de los clústeres resultantes a través de técnicas de validación.

La relevancia de este estudio radica en su capacidad para validar la implementación de las técnicas de análisis de datos en entornos empresariales reales, evidenciando como los métodos de segmentación optimizan la interpretación del comportamiento de los clientes y contribuyen a una acertada toma de decisiones estratégicas. De este modo, la investigación establece un marco metodológico replicable que puede ser empleado en futuros trabajos orientados a analizar la segmentación de clientes a través de la minería de datos.

METODOLOGÍA

Un proceso fundamental en el área de marketing es la segmentación de clientes, debido a que permite organizar la base de datos en segmentos homogéneos a partir de variables compartidas. Conforme a lo manifestado por Saxena et al. (2024) una segmentación eficaz proporciona a las empresas la capacidad de segmentar su oferta, atendiendo con

precisión las expectativas de cada grupo, acrecentando de esta manera, la satisfacción y fidelización del cliente, además de incrementar la rentabilidad empresarial.

El presente estudio aplicó un enfoque cuantitativo de carácter comparativo, orientado al análisis de datos, a través de técnicas de minería de datos para reconocer patrones de comportamiento, así como para evaluar el desempeño de técnicas de aprendizaje no supervisado en la segmentación de clientes de una Pyme comercial.

La investigación se desarrolló bajo un diseño no experimental y de corte transversal, dado que el análisis se efectuó sobre información existente, sin manipulación de variables. Se emplearon técnicas clustering, las cuales son muy utilizadas debido a que permiten agrupar en un determinado grupo las observaciones similares, y organizar en otro grupo aquellas que son diferentes (Font, 2019).

Cabe resaltar que estas herramientas son empleadas frecuentemente en estudios de análisis del comportamiento y segmentación de clientes, debido a que posibilitan la identificación tanto de tendencias como de patrones en grandes volúmenes de



información, contribuyendo no solo a la comprensión del comportamiento de los clientes, sino también al desarrollo de estrategias empresariales focalizadas y sustentadas en datos.

Para el desarrollo del estudio se empleó un conjunto de datos correspondiente a una población conformada por 200 clientes pertenecientes a una Pyme comercial, considerando como unidad de análisis el comportamiento de compra de los clientes. Con el propósito de analizar los patrones de interacción comercial se utilizaron variables vinculadas a los hábitos de compra, particularmente la frecuencia de compras, el monto de la transacción y la antigüedad del cliente.

Previo a la aplicación de las técnicas de clustering, se efectuó un preprocesamiento del conjunto de datos seleccionado para el análisis, con la finalidad de garantizar tanto su calidad como su consistencia. Este proceso comprendió la evaluación de las variables existentes, la detección de irregularidades en los datos, y la preparación de las métricas seleccionadas.

Con el objetivo de prevenir que las disparidades en las escalas de medición influyan en la formación de los

clústeres, se les aplicó a las variables, un proceso de normalización, que según Bermúdez León (2020) consiste en estandarizar la información bajo una escala de valores homogénea. De este modo, es posible comparar aquellas características que poseen distintos rangos, garantizando que ninguna de las variables domine a otra por su tamaño, con la finalidad de evitar posibles defectos que conlleven no solo a un mal diseño, sino incluso a un ineficaz procesamiento de datos, razón por la que resulta necesario aplicar una estandarización tipo z-score, empleando la siguiente fórmula:

$$Z = \frac{\chi - \mu}{\sigma}$$

La normalización permitió transformar los valores originales de las variables, a una escala común con media cero y una desviación estándar uno, lo cual permite que las variables aporten de manera equilibrada al análisis.

Este es un procedimiento que suele recomendarse con el propósito de prevenir que las variables con valores superiores, adquieran una relevancia desproporcionada, invalidando el peso relativo de las otras variables (Plazas Niño, 2021).

Para la segmentación de los datos se aplicaron dos técnicas de



agrupamiento; el clustering jerárquico y el k-means clustering. Inicialmente se implementó el clustering jerárquico aglomerativo el cual es muy popular y habitual para agrupar tanto productos como clientes similares. Este método emplea un enfoque ascendente donde cada elemento conforma inicialmente un grupo individual, posteriormente estos se agrupan progresivamente en función de sus similitudes hasta converger en un único clúster global (Oti & Olusola, 2024).

Esta técnica permite elaborar una estructura jerárquica mediante la agrupación escalonada de datos con características similares, para lo cual se empleó un dendograma para representar los resultados, posibilitando de esta manera el seguimiento gráfico de la formación de los clústeres, así como el análisis de las similitudes entre las variables, y la estimación inicial del número de agrupaciones presentes en los datos.

Posteriormente, se aplicó el K-means clustering que es un método de partición muy empleado en el aprendizaje no supervisado para dividir una base de datos en K grupos homogéneos. Su principal objetivo es reducir la suma de cuadrados intra-

clúster, a través de la asignación de cada observación al centroide más próximo. Debido tanto a su simplicidad como a su eficiencia, es aplicado frecuentemente en análisis estadísticos y minería de datos dado que aporta de manera significativa a la comprensión y exploración de arquitecturas subyacentes en conjuntos de datos complejos (Marrero et al., 2021).

Esta técnica de segmentación constituye una herramienta diseñada para fraccionar los datos a través de la minimización de la distancia entre las observaciones y el centroide de cada clúster.

Para el cálculo de la similitud entre las observaciones se aplicó la distancia euclidiana, la cual es una de las medidas más utilizadas en los análisis de clustering, debido a la simplicidad y a la fácil interpretación geométrica que posee. Esta métrica permite medir la proximidad entre las observaciones, posibilitando la formación de conjuntos homogéneos, en algoritmos como el clustering jerárquico y el K-means (Mussabayev, 2024).

La aplicación de la distancia euclidiana resulta clave tanto para la construcción de los clústeres, como para el análisis de la estructura de los datos,



dado que contribuye a la obtención de segmentos bien definidos, lo cual posibilita una toma de decisiones fundamentada en la proximidad real de los comportamientos analizados. Esta métrica se calcula mediante la siguiente fórmula:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Donde tanto x como y representan los valores de las observaciones y n equivale a la cantidad de variables consideradas en el análisis. La distancia euclidiana fue empleada en la aplicación de los clustering jerárquico y K-means, con el objetivo de definir el grado de similitud existente entre las observaciones del conjunto de datos.

Para determinar el adecuado número de clústeres, se aplicó el elbow method o método del codo, el cual es una de las técnicas más empleadas para establecer la cantidad óptima de clústeres de un conjunto de datos. Se basa en identificar visualmente el punto de inflexión donde la variación interna de los grupos se estabiliza, definiendo de este modo, el adecuado número de clústeres (Shi et al., 2021).

El método del codo se distingue como un recurso analítico estándar

ampliamente utilizado para determinar el número ideal de clústeres en los análisis de segmentación. Al ser una técnica visual clásica, posibilita la estimación previa de los segmentos en un conjunto de datos. Su principal ventaja radica en su carácter intuitivo, debido a que permite identificar el umbral exacto en el que la incorporación de más clústeres, no produce significativas mejoras en la reducción de la varianza interna.

Con el propósito de evaluar la calidad de los clústeres obtenidos, se empleó el silhouette index o índice de silueta propuesto por Rousseeuw (1987) el cual representa un indicador de verificación interna, empleado para validar la cohesión de los segmentos, permitiendo calcular el grado de similitud de una observación respecto a su propio clúster frente a los demás. Esta métrica es frecuentemente utilizada para validar los resultados de los clustering jerárquico y K-means. Para el cálculo de la silueta de una observación se emplea la siguiente fórmula:

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))}$$

Donde $a(i)$ denota la distancia promedio entre la observación i y el resto de los elementos de su propio clúster; por su parte $b(i)$ equivale a la distancia



media mínima entre la observación i y los datos del segmento diferente más cercano.

Los valores del índice de silhouette varían entre -1 y 1, los próximos a 1 denotan no solo que la observación esta asignada a su clúster de forma correcta, sino también que esta adecuadamente separada de los otros grupos; las cifras cercanas a 0 indican que la observación se sitúa en el límite entre dos clústeres, mientras que los valores con signo negativo sugieren una inadecuada asignación.

En el presente estudio, se aplicó el índice de silhouette tanto para evaluar la calidad de los grupos obtenidos, como para contrastar distintas configuraciones de la técnica k-means con diversos números de clústeres, con el propósito de determinar la segmentación óptima para los datos analizados.

Tanto el procesamiento como el análisis de los datos, se efectuó empleando el lenguaje de programación del software R, el cual es ampliamente utilizado tanto para el análisis estadístico como para la visualización de los datos. Es una herramienta desarrollada por estadísticos que proporciona un marco de trabajo integral para el riguroso análisis de datos, actualmente es uno de

los recursos más empleados en la ciencia de datos y la estadística aplicada dado no solo a la versatilidad que posee, sino también a su potencia algorítmica (Ramirez-Valverde & Ramirez-Valverde, 2022).

Esta plataforma permitió ejecutar los algoritmos de clustering, además de generar la representación gráfica de los resultados obtenidos. Para el desarrollo del estudio se utilizaron paquetes especializados como clúster que fue empleado para la implementación de las técnicas de clustering y la evaluación de la calidad de los agrupamientos; así como factoextra, el cual permitió el despliegue gráfico de los segmentos detectados.

Estos recursos facilitaron la implementación de técnicas como clustering jerárquico y k-means, además de analizar la estructura de agrupamiento disponible en los datos, mediante la generación de un dendograma, así como la consecuente detección de segmentos con características similares.

RESULTADOS

A continuación, se exponen los hallazgos obtenidos de la aplicación de las técnicas de análisis de clustering jerárquico y K-means, con el objetivo de



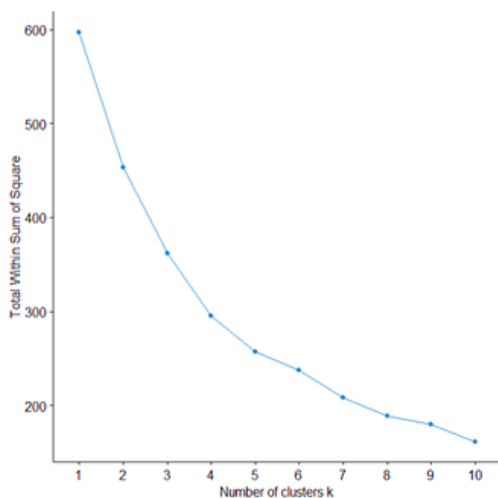
detectar patrones de segmentación, en el conjunto de datos estudiado. Además, se presentan los resultados alcanzados en la validación de los clústeres, aplicado para evaluar la calidad de los clústeres.

Determinación del número de clústeres adecuados para el análisis.

Como paso previo a la implementación del K-means clustering y con el propósito de establecer el número adecuado de clústeres para el estudio, se aplicó el método del codo, como se puede apreciar en la Figura 1. Este procedimiento posibilita evaluar la varianza intra-clúster, en función del incremento de la cantidad de clústeres considerados.

Figura 1:

Método del codo para la determinación del número óptimo de clústeres.



Fuente: Elaboración propia mediante el software R.

Como es posible observar en la Figura 1, la suma de cuadrados intra-clúster muestra una pronunciada disminución entre uno y tres clústeres, umbral donde la curva empieza a estabilizarse. Aunque la representación gráfica indica una posible solución entre dos y tres clústeres, se optó por trabajar con dos clústeres, dado que esta configuración permite una mayor claridad en la caracterización de los perfiles obtenidos, además de que facilita el análisis comparativo entre las técnicas tanto de clustering jerárquico como de k-means clustering.

Análisis de clustering jerárquico.

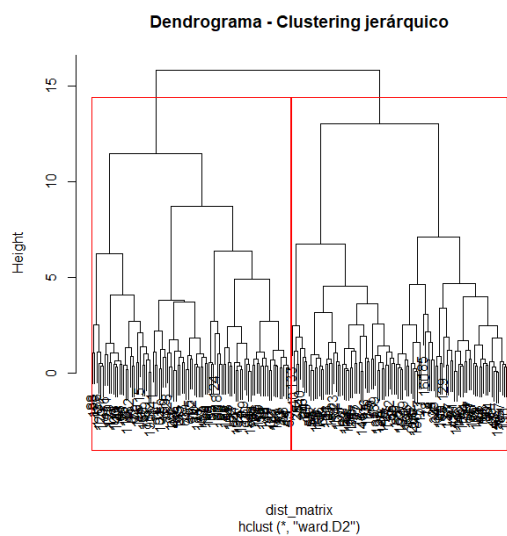
El análisis de clustering jerárquico constituye una relevante técnica exploratoria, debido a que no exige una previa definición del número de clústeres, además de que agiliza la identificación de posibles esquemas de segmentación en los datos.

Con el objetivo tanto de examinar la estructura natural de los datos, como de corroborar el número óptimo de segmentos, se aplicó la técnica de clustering jerárquico, el cual posibilita detectar agrupaciones a través de un progresivo proceso de fusión, basado en la similitud entre observaciones,



generando una representación visual mediante un dendrograma, presentado en la Figura 2.

Figura 2:
Dendrograma del clustering jerárquico



Fuente: Elaboración propia mediante el software R.

La Figura 2 muestra el dendrograma obtenido a través de la técnica del clustering jerárquico, que evidencia la dinámica de integración gradual de las observaciones según su similitud. El eje vertical representa la disimilitud o distancia existente entre los conjuntos que se fusionan, mientras que el eje horizontal refleja las unidades de análisis consideradas en el estudio.

Conforme la altura del dendrograma incrementa, las observaciones se integran en clústeres cada vez más amplios. Al trazar una línea

de corte horizontal en el nivel correspondiente, es posible identificar dos principales grupos, lo cual revela la existencia de dos segmentos claramente diferenciados dentro del conjunto de datos analizado.

Este esquema asociativo resulta consistente con la indagación exploratoria ejecutada previamente mediante el método del codo y establece la base para la implementación del algoritmo k-means clustering, que permite no solo optimizar la segmentación, sino también analizar las características de cada clúster identificado.

Para examinar las características de los clústeres detectados mediante la técnica de clustering jerárquico, se obtuvieron los promedios de las variables frecuencia, monto y antigüedad para cada segmento, los cuales se muestran en la Tabla 1. Este análisis permitió identificar aquellas variables que cuentan con una mayor capacidad discriminante en la estructuración de los conjuntos.

**Tabla 1:**

Características promedio de los clústeres obtenidos mediante el clustering jerárquico

| Clúster | Frecuencia Promedio | Monto Promedio | Antigüedad Promedio |
|----------------|----------------------------|-----------------------|----------------------------|
| 1 | 7,60 | 104,06 | 22,60 |
| 2 | 12,15 | 96,80 | 25,48 |

Fuente: Elaboración propia mediante el software R.

La Tabla 1 expone las características promedio de los clústeres obtenidos mediante la técnica de clustering jerárquico. Se observan disparidades entre los segmentos, especialmente en la variable frecuencia, en la cual el clúster 2, muestra un mayor promedio de transacciones, en contraste con el clúster 1.

En lo concerniente a la variable monto, se aprecia que el clúster 1 registra una media sutilmente superior, lo que sugeriría la existencia de observaciones con una frecuencia menor, pero de alto impacto económico. Por otro lado, el clúster 2 evidencia una alta rotación con moderados valores de transacción.

En lo que respecta a la antigüedad del cliente, se perciben disparidades moderadas entre los clústeres; no obstante, dicha variable no constituye el factor principal de diferenciación entre los segmentos. Los resultados obtenidos revelan que dentro

del conjunto de datos estudiado existen perfiles diferenciados, lo cual ratifica la capacidad del clustering jerárquico para identificar relevantes estructuras de segmentación.

Análisis de K-means clustering

Con el propósito de optimizar la segmentación identificada en el análisis jerárquico, se implementó el algoritmo k-means clustering, mismo que permite fraccionar las observaciones, en un número predefinido de clústeres, disminuyendo la dispersión intra-clúster y potenciando la distancia entre ellos. Cabe resaltar que esta técnica es empleada ampliamente en los análisis de segmentación debido no solo a su alto rendimiento algorítmico, sino también a la capacidad que posee para detectar patrones homogéneos en grandes conjuntos de datos.

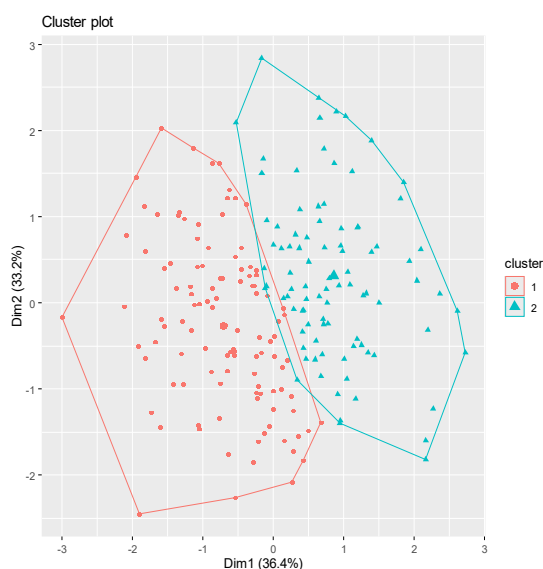
Con base en los resultados obtenidos a través del método del codo y el clustering jerárquico, se estableció un



total de dos clústeres. Se implementó el análisis mediante el software estadístico **R**, empleando datos estandarizados previamente, con la finalidad de mitigar sesgos procedentes de las diferencias de escala entre las variables. La Figura 3 muestra la segmentación obtenida mediante la aplicación del k-means.

Figura 3:

Visualización de los clústeres obtenidos mediante K-means.



Fuente: Elaboración propia mediante el software R.

En la Figura 3 se muestran los resultados derivados del algoritmo k-means clustering, en el cual se distinguen dos clústeres claramente diferenciados. Cada uno de los puntos corresponde a una observación del conjunto de datos, por su parte los colores rojo y azul denotan la pertenencia a cada segmento.

Se aprecia una separación adecuada entre los clústeres, especialmente en función de la primera dimensión (Dim 1), lo que refleja la presencia de desigualdades estructurales entre las agrupaciones identificadas. Además, los polígonos que bordean cada segmento de observaciones posibilitan representar la variabilidad intra-clúster, y evidencian una asociación relativamente consistente dentro de cada grupo.

La representación gráfica de los datos en dos dimensiones, representa a un proceso de reducción de dimensionalidad aplicada con propósitos de visualización, conservando la estructura de los clústeres generada por el algoritmo. Los resultados obtenidos corroboran la presencia de dos principales grupos dentro del conjunto de datos, validando los hallazgos alcanzados a través del clustering jerárquico.

Con el propósito de caracterizar de forma cuantitativa, los clústeres obtenidos mediante el K-means clustering, se calcularon los promedios de las variables frecuencia, monto y antigüedad para cada uno de los segmentos, los cuales se muestran en la Tabla 2. Dichos resultados permiten



reconocer las métricas con mayor potencial discriminante en la estructuración de los conjuntos.

Tabla 2:

Características promedio de los clústeres obtenidos mediante el K-means clustering.

| Clúster | Frecuencia Promedio | Monto Promedio | Antigüedad Promedio |
|----------------|----------------------------|-----------------------|----------------------------|
| 1 | 7,83 | 109,04 | 23,83 |
| 2 | 12,11 | 90,96 | 24,16 |

Fuente: Elaboración propia mediante el software R.

La Tabla 2 expone las características promedio de los grupos identificados a través del K-means clustering, reflejando disparidades significativas en las variables frecuencia y monto promedio de las transacciones. El clúster 1 presenta una frecuencia de transacciones menor, pero con un monto promedio alto, lo cual indica la presencia de clientes con una baja recurrencia, aunque con compras de mayor valor. Por el contrario, el segmento 2 muestra una alta frecuencia de transacciones, y un monto promedio bajo, lo que denota un perfil de clientes con un mayor dinamismo comercial, pese a que los valores de compra son moderados.

La variable antigüedad, por su parte, evidencia valores similares en los dos clústeres, lo que indica que no representa un factor crítico en la

diferenciación de los grupos. Bajo esta perspectiva las disparidades detectadas sugieren claramente que las variables frecuencia y monto poseen una mayor capacidad discriminante en la estructuración de los clústeres.

Los resultados, en conjunto, manifiestan que la segmentación lograda permite detectar perfiles de comportamiento diferenciados, lo cual confirma la efectividad del K-means clustering como herramienta tanto para el análisis como para la clasificación de los clientes en función de sus patrones de compra.

Los valores promedios obtenidos permiten reconocer patrones de comportamiento diferenciados entre los clústeres. El clúster 1 muestra una frecuencia de transacciones baja, pero con un monto promedio mayor, lo cual



sugiere un segmento de clientes con una interacción comercial esporádica, por lo que podrían ser considerados como clientes ocasionales.

El clúster 2 por su parte, presenta una alta frecuencia de transacciones con montos promedio moderados, lo que refleja un grupo de clientes con un comportamiento de consumo más recurrente, por lo que podría perfilarse como clientes frecuentes.

Esta caracterización facilita una mayor comprensión de los grupos detectados, así como de los patrones de comportamiento presentes en el conjunto de datos analizado. Además, los resultados obtenidos ponen de manifiesto la eficacia de la técnica K-

means para detectar perfiles diferenciados, lo que respalda su relevancia como herramienta estratégica tanto para la segmentación como para el estudio del comportamiento de las observaciones.

Comparación de las técnicas de clustering aplicadas.

Con el objetivo de contrastar la fiabilidad de los resultados alcanzados a través del clustering jerárquico y el algoritmo k-means, se realizó un análisis comparativo de la distribución de las observaciones entre ambas técnicas, el cual se muestra en la Tabla 3. Este procedimiento permite determinar el nivel de concordancia en la segmentación obtenida.

Tabla 3:

Análisis comparativo de la distribución de los clústeres entre clustering jerárquico y K-means

| Clustering Jerárquico | K-means C1 | K-means C2 |
|------------------------------|-------------------|-------------------|
| C1 | 93 | 11 |
| C2 | 16 | 80 |

Fuente: Elaboración propia mediante el software R.



De acuerdo con los datos expuestos en la Tabla 3, se manifiesta una elevada concordancia entre los resultados obtenidos mediante el clustering jerárquico y el K-means. Específicamente, 93 observaciones que el método jerárquico clasificó en el clúster 1, fueron también asignadas por k-means al mismo grupo, mientras que 80 observaciones del segundo clúster presentan una clasificación idéntica en ambos procedimientos.

A nivel general, 173 de los 200 datos analizados, que equivale al 86,5% de las observaciones presentaron una asignación concordante por ambos clusterings, lo cual refleja una estabilidad alta en la estructura de segmentación identificada. Las discrepancias registradas en 27 casos responden a las características metodológicas de cada técnica, dado que el clustering jerárquico se rige por un esquema de proximidad sucesiva, por su parte el k-means optimiza la varianza intra-clúster.

Los hallazgos obtenidos permiten inferir que ambas técnicas convergen en la detección de patrones similares dentro del conjunto de datos, lo que fortalece la validez de la segmentación alcanzada, además de que

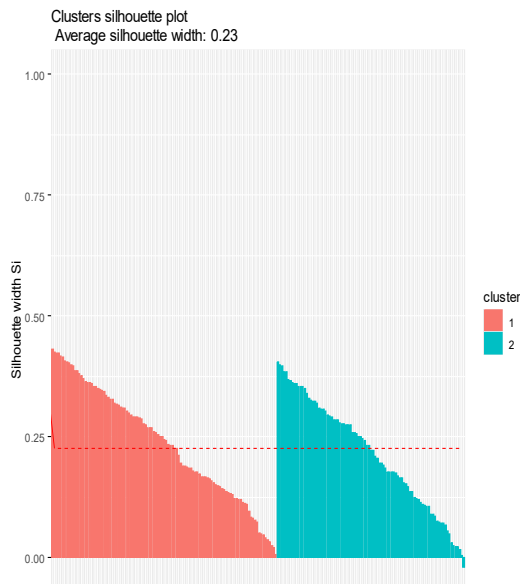
ratifica la presencia de perfiles diferenciados definidos claramente. En este sentido, la estabilidad detectada refuerza la fiabilidad de los resultados y avala la implementación combinada de ambos métodos para el análisis de segmentación.

Validación de los clústeres

Como parte del proceso de validación de la segmentación obtenida, se implementó el índice de silhouette, que se muestra en la Figura 4, con el propósito de analizar tanto la calidad, como la consistencia de la agrupación generada por el K-means clustering. Este indicador permite calcular el grado de similitud de cada dato con su clúster de pertenencia, en comparación con los demás grupos, facilitando una estimación de la cohesión interna y la distancia entre los segmentos. Debido a estas particularidades, el índice de silhouette, se consolida como uno de los parámetros más sólidos y frecuentemente empleados para validar la calidad de los hallazgos en los análisis de clustering.



Figura 4:
Representación gráfica del índice de silhouette para la validación de los clústeres obtenidos con K-means



Fuente: Elaboración propia mediante el software R.

Como se aprecia en la Figura 4, el valor promedio del índice de silhouette es 0,226 (0,23 en la representación gráfica), lo cual denota una aceptable calidad de agrupamiento. Las observaciones, en su mayoría, muestran valores positivos, lo que sugiere que han sido asignadas correctamente a sus respectivos clústeres, y que existe una homogeneidad intra-clúster adecuada en los segmentos detectados.

Pese a ello, se identifican ciertas observaciones con valores próximos a cero, lo cual refleja una ligera superposición entre los segmentos. Este

fenómeno resulta previsible en análisis de segmentación de carácter exploratorio, donde los límites inter-clúster carecen de una demarcación clara.

Los hallazgos corroboran que los clústeres obtenidos mediante k-means presentan una estructura equilibrada, lo cual avala la funcionalidad del modelo para el reconocimiento de patrones dentro del conjunto de datos estudiado. Además, el elevado nivel de coincidencia observado entre los resultados tanto del clustering jerárquico como del k-means, potencia la confiabilidad y estabilidad de la segmentación lograda.

En síntesis, los resultados obtenidos mediante la implementación del método del codo, el clustering jerárquico y K-means, revelan la existencia de una consistente estructura de segmentación al interior del conjunto de datos estudiados. La alta similitud en la asignación de observaciones entre ambos métodos, indica que los segmentos identificados cuentan con características diferenciadas claramente establecidas.

De igual manera, la validación realizada mediante el índice de silhouette, que obtuvo un valor de 0,226,



ratifica que la segmentación es de calidad aceptable, respaldando así la idoneidad de los modelos empleados para la identificación de patrones en los datos.

En este contexto, los resultados permiten afirmar que las técnicas de clustering aplicadas, son herramientas adecuadas para la segmentación del conjunto de datos, permitiendo identificar perfiles con comportamientos variados.

DISCUSIÓN

En el ámbito de la segmentación de clientes y el análisis de clústeres, se han aplicado distintos métodos para gestionar los desafíos de los datos mixtos y los entornos de información a gran escala. Los resultados obtenidos en el presente estudio demuestran que el algoritmo K-means y el procedimiento de clustering jerárquico son capaces de detectar estructuras de segmentación coherentes en el conjunto de datos evaluado.

La similitud apreciada en la distribución de las observaciones entre ambas técnicas, indica que los datos poseen patrones relativamente estables, lo cual avala la idoneidad del uso combinado de estos métodos para

potenciar la fiabilidad del análisis de segmentación.

Desde un punto de vista metodológico, el clustering jerárquico, posibilitó detectar de forma exploratoria una posible estructura de los datos a través de un análisis de proximidad entre las observaciones. Este tipo de métodos sigue siendo relevante para analizar datos complejos, dado a la capacidad que tiene para gestionar estructuras heterogéneas y facilitar tanto la interpretación como la comprensión de los patrones de cluster identificados (Han et al., 2024).

Por su parte el algoritmo K-means permitió optimizar la segmentación de los datos, mediante la disminución de la variabilidad intra-clúster. Este patrón es congruente con investigaciones previas, las cuales resaltan la eficiencia de K-means para el agrupamiento de datos, la identificación de patrones y la segmentación de clientes, facilitando de esta manera la comprensión tanto de las preferencias como del comportamiento de los usuarios, sustentando no solo la toma de decisiones, sino también el diseño de estrategias eficientes (Naoui et al., 2020).



En lo que respecta a la validación del modelo, el resultado de 0,226 obtenido mediante el índice silhouette, sugiere que la calidad de la segmentación es aceptable. De acuerdo con (Chicco et al., 2025) lograr valores positivos en el índice silhouette confirma que las observaciones han sido asignadas correctamente a sus respectivos clústeres, aun cuando la separación entre ellos no sea completamente definida. Dicha situación es recurrente en investigaciones de carácter empírico, dado que, al trabajar con datos reales, es común que distintos perfiles compartan algunos intereses o comportamientos.

Asimismo, la caracterización de los clústeres permitió la identificación de características distintivas entre los conjuntos analizados, especialmente en relación con la frecuencia, monto y antigüedad; aquello evidencia la capacidad que tienen las técnicas de clustering para facilitar la comprensión de patrones de consumo presente en los datos. Los resultados obtenidos concuerdan con lo manifestado por Tan et al. (2019) quienes resaltan que los métodos de segmentación posibilitan la simplificación de la complejidad de los datos a través de la detección de estructuras homogéneas.

Sin embargo, la presente investigación reconoce ciertas limitaciones; en primera instancia, el análisis se sustentó en un determinado conjunto de variables, por ende la adhesión de nuevas características podría causar cambios en la estructura de los clústeres identificados; asimismo, el análisis comparativo se restringió a dos técnicas de clustering, por lo que se sugiere que estudios futuros podrían acrecentar la comparación al incorporar otros algoritmos de segmentación para reforzar la validación de los resultados.

Como propuesta para futuros trabajos investigativos, resulta conveniente examinar la estabilidad de los clústeres a través de distintas métricas de validación tanto interna como externa: se sugiere además estudiar el comportamiento de estas técnicas en grandes conjuntos de datos o con variables suplementarias que aporten una mayor profundidad al análisis de segmentación.

CONCLUSIONES

El trabajo desarrollado posibilitó la identificación de la presencia de patrones diferenciados, al interior del conjunto de datos analizado a través de la implementación de técnicas de clustering jerárquico y K-means,



ratificando la eficacia de estas herramientas tanto para el análisis exploratorio, como para la segmentación de datos multivariados.

Los hallazgos revelaron que ambas técnicas posibilitan detectar estructuras de agrupación coherentes, apreciándose un elevado grado de convergencia en las clasificaciones realizadas por cada clustering. Aquello indica que los métodos empleados constituyen herramientas complementarias que consolidan la fiabilidad y consistencia del proceso de segmentación.

Adicionalmente, la caracterización de los clústeres, permitió detectar perfiles diferenciados, según variables como frecuencia, monto y antigüedad, esto propició un entendimiento más profundo de los conjuntos identificados, además de que demostró el potencial de estos métodos para reducir la complejidad informativa de los datos.

Desde una perspectiva metodológica, el empleo integrado del clustering jerárquico como técnica exploratoria y K-means como método de optimización, facilitó la consecución de una segmentación más sólida, demostrando la relevancia de emplear

enfoques comparativos en estudios de análisis de datos.

En este contexto, la investigación proporciona sustento empírico sobre la idoneidad de las técnicas de aprendizaje no supervisado para el reconocimiento de estructuras de segmentación en conjuntos de registros, aportando a reforzar la implementación de herramientas de minería de datos en procesos de análisis estadístico.

Por último, los resultados obtenidos corroboran que la implementación de técnicas de clustering representa una estrategia eficaz para detectar patrones y reducir la complejidad de los datos, posibilitando la generación de información relevante para comprender el comportamiento de las observaciones analizadas.

De este modo, el estudio refleja la relevancia de aplicar métodos de análisis multivariado para analizar datos complejos, resaltando la eficacia del clustering para la detección de estructuras internas que resultan imperceptibles a través de análisis tradicionales.

En este sentido la investigación consolida la importancia de emplear técnicas de minería de datos como soporte en los procesos de análisis



estadístico y la toma de decisiones fundamentadas en datos, proporcionando una perspectiva metodológica que puede ser replicada en estudios futuros que requieran la segmentación de información.

REFERENCIAS

- Abdel-Hakim, A. E., Ibrahim, A.-M. M., Bouazza, K. E., Deabes, W., & Hedar, A.-R. (2024). Ellipsoidal K-Means: An Automatic Clustering Approach for Non-Uniform Data Distributions. *Algorithms*, 17(12), 551. <https://doi.org/https://doi.org/10.3390/a17120551>
- Afzal, A., Khan, L., Hussain, M. Z., Hasan, M. Z., Mustafa, M., Khalid, A., . . . Javaid, A. (2024). Customer Segmentation Using Hierarchical Clustering. 2024 *IEEE 9th International Conference for Convergence in Technology (I2CT)* (págs. 1-6). Pune, India: IEEE. <https://doi.org/https://doi.org/10.1109/I2CT61223.2024.10543349>
- Alonso, J., Largo, M., & Hoyos, C. (2024). *Una introducción a los modelos de Clustering empleando R*. Universidad Icesi. <https://doi.org/https://doi.org/10.18046/EUI/bda.h.6>
- Alves Gomes, M., & Meisen, T. (2023). A review on customer segmentation methods for personalized customer targeting in e-commerce use cases. *Information Systems and e-Business Management*, 21, 527-570. <https://doi.org/https://doi.org/10.1007/s10257-023-00640-4%0A>
- Ayala-Aldana, N., & Parra-Cid, C. (2023). Algoritmos de agrupamiento jerárquico para el control de la susceptibilidad antibiótica. *Revista Chilena de Enfermedades Respiratorias*, 39(1), 120-121. Obtenido de <https://revchilenfermrespir.cl/index.php/RChER/article/view/1124>
- Bermúdez León, M. (2020). *Normalización*. Universidad San Marcos. Obtenido de <https://repositorio.usam.ac.cr/xmlui/bitstream/handle/11506/2125/LEC%20ING%20SIST%200069%202020.pdf?sequence=1&isAllowed=y>



- Camargo Morales, F., Angarita López, J., & Najjar Sánchez, O. (2023). *Sistematización de conceptos de marketing con realidad aumentada*. Ediciones de la U. Obtenido de https://books.google.com.ec/books?id=69jDEAAAQBAJ&pg=PA88&hl=es&source=gbs_selected_pages&cad=1#v=onepage&q&f=true
- Chicco, D., Campagner, A., Spagnolo, A., Ciucci, D., & Giuseppe, J. (2025). The Silhouette coefficient and the Davies-Bouldin index are more informative than Dunn index, Calinski-Harabasz index, Shannon entropy, and Gap statistic for unsupervised clustering internal evaluation of two convex clusters. *PeerJ Computer Science*, *11*, e3309. <https://doi.org/10.7717/peerj-cs.3309>
- Font, X. (2019). *Técnicas de clustering*. FUOC. Obtenido de <https://openaccess.uoc.edu/server/api/core/bitstreams/859ca353-d4f7-4448-a284-6454decfc950/content>
- Franco-Árcega, A., Sobrevilla-Sólis, V., Gutiérrez-Sánchez, M., García-Islas, L., Suárez-Navarrete, A., & Rueda-Soriano, E. (2021). Sistema de enseñanza para la técnica de agrupamiento k-means. *Padi Boletín Científico de Ciencias Básicas e Ingenierías del ICBI*, *9*(Especial), 53-58. <https://doi.org/https://doi.org/10.29057/icbi.v9iEspecial.7384>
- Guañuna Viteri, N. (2022). *Clasificación de los productos de una empresa de Quito considerando el recurso tiempo de mano de obra asignado a cada uno de sus procesos durante el año 2021-2022 utilizando algoritmos de aprendizaje no supervisado*. Obtenido de [Tesis de maestría, Pontificia Universidad Católica del Ecuador]: <https://repositorio.puce.edu.ec/server/api/core/bitstreams/268b664d-220d-468b-a8d3-a68d5469976f/content>
- Han, W., Zhang, S., Gao, H., & Bu, D. (2024). Clustering on hierarchical heterogeneous data with prior pairwise relationships. *BMC Bioinformatics*, *25*(40), 1-22.



- <https://doi.org/https://doi.org/10.1186/s12859-024-05652-6>
- Kopřivová, V., & Matušínská, K. (2023). Unlocking Generation Y: Market Segmentation via Lifestyle Insights. *Communication Today*, 14(2), 122-139. <https://doi.org/https://doi.org/10.34135/communicationtoday.2023.Vol.14.No.2.9>
- Marrero, L., Carrizo, D., García-Santander, L., & Ulloa-Vasquez, F. (2021). Uso de algoritmo K-means para clasificar perfiles de clientes con datos de medidores inteligentes de consumo eléctrico: Un caso de estudio. *Ingeniare. Revista chilena de ingeniería*, 29(4), 778-787. <https://doi.org/http://dx.doi.org/10.4067/S0718-33052021000400778>
- Moharana, U., & Sarmah, S. (2018). Joint replenishment of associated spare parts using clustering approach. *The International Journal of Advanced Manufacturing Technology*, 94(1), 2535-2549. <https://doi.org/https://doi.org/10.1007/s00170-017-0909-6>
- Mussabayev, R. (2024). Optimizing Euclidean Distance Computation. *Mathematics*, 12(23), 3787. <https://doi.org/https://doi.org/10.3390/math12233787>
- Naoui, M. A., Brahim, L., & Mouloud, A. (2020). Usando el algoritmo K-means para la curva de regresión en un gran sistema de datos para el entorno empresarial. *Revista Cubana de Ciencias Informáticas*, 14(2), 34-48. Obtenido de <https://rcci.uci.cu/index.php/RC-CI/article/view/1909>
- Navarrete Vines, M., Alegría Camino, D., Galarza Luna, A., & Ramírez Garofalo, D. (2025). La Inteligencia Artificial y el aprendizaje automático en la Educación Superior del Ecuador. *Tesla Revista Científica*, 5(2), e515. <https://doi.org/https://doi.org/10.55204/trc.v5i2.e515>
- Oti, E. U., & Olusola, M. O. (2024). Overview of Agglomerative Hierarchical Clustering Methods. *British Journal of Computer, Networking and Information Technology*, 7(2), 14-23.



- <https://doi.org/https://www.doi.org/10.52589/BJCNIT-CV9POOGW>
- Plazas Niño, F. (2021). *Introducción al análisis clúster: Una aplicación en la clasificación de campos petroleros*. Obtenido de https://www.researchgate.net/publication/350936275_INTRODUCCION_AL_ANALISIS_CLUSTER_UNA_APLICACION_EN_LA_CLASIFICACION_DE_CAMPOS_PETROLEROS
- Prieto Herrera, J. E. (2021). *Investigación de mercados* (Tercera ed.). ECOE. Obtenido de https://books.google.com.ec/books?id=tQpZEAAAQBAJ&printsec=frontcover&source=gbs_atb&redir_esc=y#v=onepage&q&f=false
- Ramírez-Valverde, G., & Ramírez-Valverde, B. (2022). Programa estadístico R, Herramienta clave en el análisis y visualización de datos. *Agro-Divulgación*, 2(2), 17-22. Obtenido de <https://agrodivulgacion-colpos.org/index.php/1agrodivulgacion1/article/view/59>
- Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53-65. [https://doi.org/https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/https://doi.org/10.1016/0377-0427(87)90125-7)
- Saxena, A., Agarwal, A., Pandey, K. B., & Pandey, D. (2024). Examination of the Criticality of Customer Segmentation Using Unsupervised Learning Methods. *Circular Economy and Sustainability*, 4, 1447-1460. <https://doi.org/https://doi.org/10.1007/s43615-023-00336-4>
- Shi, C., Wei, B., & Wei, S. (2021). A quantitative discriminant method of elbow point for the optimal number of cluster in clustering algorithm. *EURASIP Journal on Wireless Communications and Networking*, 2021(31). <https://doi.org/https://doi.org/10.1186/s13638-021-01910-w>
- Singh, M., Singh, A., Gupta, M., & Reddy, R. (2022). Leveraging K-Means Clustering and Hierarchical Agglomerative Algorithms for Scalable AI-Driven Customer Segmentation. *Journal of AI ML Research*,



- 11(7), 1-23. Obtenido de <https://www.joaimlr.com/index.php/v1/article/view/31>
- Stack Sánchez, P. (2021). *Métodos de aprendizaje supervisado y no supervisado para la estimación de microestructura cerebral en datos de DWMR*. Obtenido de [Tesis de maestría, Centro de Investigación en Matemáticas, A.C.]: <https://cimat.repositorioinstitucional.mx/jspui/bitstream/1008/1129/1/TE%20835.pdf>
- Tabianan, K., Velu, S., & Ravi, V. (2022). K-Means Clustering Approach for Intelligent Customer Segmentation Using Customer Purchase Behavior Data. *Sustainability*, 14(12), 7243. <https://doi.org/https://doi.org/10.3390/su14127243>
- Tan, P.-N., Steinbach, M., Karpatne, A., & Kumar, V. (2019). *Introduction to Data Mining*. Pearson Education. Obtenido de https://api.pageplace.de/preview/DT0400.9780273775324_A37747616/preview-9780273775324_A37747616.pdf
- Trujillo Logroño, K. (2025). *Análisis del proceso de clustering con K-means para la segmentación de clientes en la empresa de Rastreo 365*. Obtenido de [Tesis de maestría, Universidad Tecnológica Indoamérica]: <https://repositorio.uti.edu.ec/server/api/core/bitstreams/a89f526c-234c-473a-ba98-0ee2aff815c7/content>
- Ufeli, C., Sattar, M., Hasan, R., & Mahmood, S. (2025). Enhancing Customer Segmentation Through Factor Analysis of Mixed Data (FAMD)-Based Approach Using K-Means and Hierarchical Clustering Algorithms. *information*, 16(6), 441. <https://doi.org/https://doi.org/10.3390/info16060441>
- Uvidia Fassler, M., Cisneros Barahona, A., Méndez Naranjo, P., & Villa Yáñez, H. (2018). Minería de datos para la toma de decisiones en la unidad de nivelación y admisión universitaria ecuatoriana. *Revista Cumbres*, 4(2), 55-67. Obtenido de <https://revistas.utmachala.edu.ec/revistas/index.php/Cumbres/article/view/286/127>



Wani, A. A. (2024). Comprehensive analysis of clustering algorithms: exploring limitations and innovative solutions. *PeerJ Computer Science*, 10, e2286.

<https://doi.org/https://doi.org/10.7717/peerj-cs.2286>