



Capítulo 5.

El análisis de clúster como herramienta para la toma de decisiones basada en la tipificación de los usuarios

Holmes Rodríguez-Espinosa

Ing. Agric, MSc, PhD, Profesor titular, Facultad de Ciencias Agrarias Universidad de Antioquia, Grupo de Investigación Gamma

Luis Fernando Restrepo Betancur

Estad., Esp. Profesor titular de la Facultad de Ciencias Agrarias, Universidad de Antioquia, Grupo Statistical

Diana Guzmán Álvarez

Ing. Agropec., DrSc(c). Profesora de cátedra, Universidad de Antioquia, Grupo de Investigación Gamma

Resumen

El análisis de clúster es una técnica utilizada para agrupar un conjunto de datos en grupos (clústeres) con relativa homogeneidad, si bien son lo más heterogéneos entre ellos. Esta técnica, ampliamente utilizada en la literatura científica en tanto herramienta útil para la toma de decisiones, ha sido poco utilizada en los procesos de planificación rural.

El objetivo de este estudio es proponer una metodología apoyada en el *clustering* para clasificar usuarios por medio de la identificación de las relaciones entre sus características, con el fin de mejorar los procesos de planificación en el sector agropecuario. Se realizó un estudio de tipo cualitativo exploratorio empleando la técnica de estudio de caso sobre análisis de preferencias por actividades e infraestructura disponibles en el destino turístico donde se realizó un análisis de clúster. La metodología propuesta permitió clasificar los individuos en 3 clústeres e identificar la heterogeneidad entre sus preferencias y las características sociodemográficas de los individuos en cada conglomerado. Estos aspectos facilitan la toma de decisiones sobre estrategias de atención para cada agrupamiento con base en sus características particulares. Se concluye que el análisis de clúster es una herramienta con gran potencial para mejorar los procesos de planificación en el sector agropecuario a partir de la clasificación de usuarios con base en sus características específicas.

Palabras clave: *análisis de conglomerados, análisis multivariante, estrategias de desarrollo rural, planificación agrícola, planificación rural*

I. Introducción

La clasificación de los usuarios con base en su perfil es muy utilizada en la actualidad, en particular en las redes sociales, para generar contenidos personalizados en función de las preferencias de cada usuario, lo que genera mayores niveles de satisfacción. Una de las técnicas más utilizadas para el mejoramiento de la personalización de los contenidos es la modelación del comportamiento (Knijnenburg et al., 2012) por medio de modelos automatizados de análisis de información, para lo cual

se definen las variables dependientes e independientes que permitan obtener grupos homogéneos al disminuir la varianza entre individuos y maximizar la varianza entre grupos.

El análisis de clúster es una de las técnicas más utilizadas para la categorización de grupos. Es una técnica estadística multivariante, conocida también como análisis de conglomerado, que busca agrupar un conjunto de datos en grupos con relativa homogeneidad, que sean lo más heterogéneos entre ellos, los cuales se denominan clúster (de la Fuente, 2011). Se emplean 4 procedimientos de generación de conglomerados: 1) jerárquicos; 2) basados en densidad; 3) basados en redes y 4) particionamiento. Los algoritmos jerárquicos utilizan dos métodos: la aglomeración repetida, en la que se fusionan dos clústeres, y la división repetidamente, en la cual el clúster se parte en dos. Los algoritmos basados en densidad generan clústeres de tamaño arbitrario y con valores extremos; los algoritmos basados en redes permiten el agrupamiento eficiente de datos de gran tamaño (de la Fuente, 2011).

La literatura científica sobre categorización de individuos es amplia y se destacan experiencias como clasificación de usuarios de servicios públicos (Carpaneto et al., 2006), servicios turísticos (Ye et al., 2009), lectores de un periódico digital —a través de la minería de datos— (de la Hoz et al., 2017), imágenes satelitales —por medio de redes neuronales— (Arista-Jalife et al., 2017), programas de ingeniería industrial —que utilizan aprendizaje automático— (Fontalvo-Herrera et al., 2018) y perfil del turista para descubrir grupos de consumidores homogéneos y acondicionar la oferta a sus preferencias y demandas (Molina, 2007).

Para una clusterización adecuada es fundamental establecer la distancia o disimilitud de los individuos teniendo en cuenta el ángulo y la magnitud; entre las distancias más utilizadas se encuentran la distancia

euclidiana y Bray-Curtis. Estas permiten predecir la pertenencia de un individuo a un conglomerado cuando se cuenta con la información de interés para la investigación.

A partir de lo anterior y teniendo en cuenta la aplicabilidad de esta técnica en la agricultura, el objetivo de este estudio fue proponer una metodología para la toma de decisiones en el sector agropecuario apoyada en el *clustering* que permitiera la clasificación de usuarios por medio de la identificación de las relaciones entre sus preferencias.

2. Metodología

Se realizó un estudio de tipo cualitativo exploratorio multidimensional de carácter transversal empleando la técnica de estudio de caso, con el propósito de proponer una metodología para la toma de decisiones aplicable al sector agropecuario con base en el análisis de clúster. Este estudio se realizó en tres fases: 1) revisión de literatura para la identificación del marco conceptual y estado del arte sobre análisis de clúster; 2) diseño del modelo de toma de decisiones con enfoque de clúster y 3) aplicación del modelo a un caso de segmentación de usuarios.

La primera fase se realizó por medio de un estudio interpretativo (Lamnek, 2005), a través de la técnica de análisis temático de contenido (Krippendorff, 2004), para la identificación de los hallazgos más importantes reportados en la literatura científica sobre el análisis de clúster. Se partió del establecimiento de los criterios de búsqueda para la identificación de las publicaciones requeridas (Rudas et al., 2013).

Posteriormente se realizó la búsqueda en bases de datos especializadas como Science Direct, Scielo y Dialnet, en las que se utilizaron como términos clave: “análisis de clúster”, “agricultura”, “planificación” y sus

equivalentes en inglés. También se hizo el análisis preliminar de los documentos identificados para establecer su pertinencia para el estudio, proceso en el cual se obtuvo un total de 43 artículos científicos.

Para la segunda fase se realizó el diseño del modelo de toma de decisiones con enfoque de clúster, compuesto por cuatro etapas: 1) selección de las variables clave, 2) definición del método de análisis de clúster, 3) selección del número de conglomerados y 4) análisis de las características de cada conglomerado.

En la tercera fase se realizó la aplicación del modelo para la toma de decisiones a un caso de segmentación de usuarios de agroecoturismo, como parte del proceso de validación de este, para ilustrar la forma de utilización.

3. Resultados y discusión

3.1. Marco conceptual

3.1.1. Análisis de conglomerados

Es un método exploratorio de tipo multidimensional que consiste en la identificación de unidades u objetos similares entre sí para la conformación de subconjuntos homogéneos al interior de cada grupo con diferencia entre ellos, lo cual facilita la interpretación del conjunto de datos; estos conjuntos se identifican por medio de técnicas estadísticas que tienen en cuenta el concepto de distancias geométricas de carácter multivariante. Esto implica: seleccionar, y en algunas ocasiones estandarizar, las variables relevantes para identificar los grupos, elegir la medida más adecuada de proximidad entre los individuos, seleccionar el método de conformación de los conglomerados y definir los criterios de evaluación para la identificación del número óptimo de grupos (Amicarelli et al., 2021).

3.1.2. Clúster

Se refiere al grupo de individuos o casos que resultan del análisis de conglomerados correspondiente a un conjunto de variables de naturaleza cuantitativa, de tipo continuo o discreto, o cualitativa, de tipo nominal u ordinal, agrupadas por su homogeneidad (Amicarelli et al., 2021).

3.1.3. Algoritmos para el análisis de clúster

Los algoritmos determinan las diferentes formas de medir la distancia entre clústeres, la cual genera diferentes agrupaciones. Esta es una decisión subjetiva del investigador, pues no se registra un criterio para seleccionar cuál es el mejor algoritmo; en este sentido, la decisión depende del método que refleje de mejor manera el propósito del estudio (de la Fuente, 2011).

3.1.4. Tipos de análisis de clúster

Existen dos tipos de análisis de clúster: no jerárquicos y jerárquicos. En el primero los datos se asignan a grupos que el propio análisis configura sin que unos dependan de otros, es decir, se forman grupos homogéneos sin establecer relaciones entre ellos. En el segundo se utilizan estructuras arborescentes en las que los clústeres más bajos son englobados por los de niveles superiores, es decir, los grupos se van fusionando de manera sucesiva siguiendo una jerarquía, de forma que la homogeneidad va decreciendo conforme se hacen más amplios los grupos (de la Fuente, 2011).

3.1.5. Métodos jerárquicos

Se utilizan para conformar un nuevo grupo, o separar uno existente, al maximizar la medida de similitud o minimizar la distancia. Los métodos jerárquicos se clasifican en asociativos o aglomerativos y disociativos: los primeros parten de tantos grupos como casos se tienen, los cuales se van agrupando hasta conformar grupos homogéneos; los disociativos

parten de un solo grupo con todos los casos, lo cual permite conformar grupos más pequeños a través de divisiones sucesivas (de la Fuente, 2011).

Los métodos jerárquicos asociativos más conocidos son: estrategia de la distancia mínima o similitud máxima, estrategia de la distancia máxima o similitud mínima, estrategia de la distancia o similitud promedio ponderada, métodos basados en el centroide, fórmula de recurrencia de Lance y Williams (método no ponderado), método de la mediana y método de Ward. Los métodos jerárquicos disociativos más comunes son: Linkage simple, Linkage completo, promedio entre grupos, método del centroide, método de la mediana, método de Ward y análisis de asociación (de la Fuente, 2011).

3.1.6. Método no jerárquico

Se utiliza para la clasificación de individuos, no de variables, definiendo la cantidad deseada de grupos a partir de los cuales se realiza el intercambio de los miembros de cada clúster hasta obtener la mejor agrupación. Los métodos no jerárquicos se clasifican en: reasignación, búsqueda de densidad, directos y reductivos. Los métodos jerárquicos de reasignación son k-medias y nubes dinámicas; los de búsqueda de densidad son análisis modal, métodos taxap, método de Fortin y método de Wolf; los métodos directos son block y clustering y, por último, el método reductivo es análisis factorial tipo Q (de la Fuente, 2011).

3.1.7. Distancia entre conglomerados

Se refiere a la distancia entre las observaciones y se pueden determinar de diversas formas: vecino más cercano, vecino más lejano, promedio de grupo y centroide (de la Fuente, 2011). El vecino más cercano se aplica cuando los conglomerados son más alargados y se tiene en cuenta la distancia entre los individuos más próximos de los conglomerados; el vecino más lejano se aplica cuando los conglomerados son más esféricos

y se tiene en cuenta la distancia entre los individuos más alejados de los conglomerados. Por su parte, el promedio de grupos se aplica cuando los conglomerados son más robustos, para ello se promedia la distancia entre los individuos dentro de estos grupos. Finalmente, el centroide se aplica para conglomerados más robustos al considerar la distancia entre los centroides o centros de gravedad de los conglomerados.

3.1.8. Medida de proximidad

La proximidad entre variables o individuos refleja el grado de asociación o relación por medios numéricos que caracterizan las relaciones entre ellos en un sentido específico, por lo que se debe elegir la medida apropiada para la situación concreta que se pretende estudiar. La medida de asociación puede ser distancia o similitud. La distancia permite conformar grupos con individuos parecidos en los que la distancia es pequeña entre ellos; una de las más utilizadas es la distancia euclídea. La similitud, por otra parte, permite conformar grupos con individuos con una semejanza alta entre ellos; algunas de las más utilizadas son la correlación de Pearson y los coeficientes de Spearman y de Kendall (de la Fuente, 2011).

3.1.9. Dendograma

Es una representación gráfica del proceso de agrupación que resulta del análisis de clúster, en la cual los individuos o variables similares se conectan por medio de enlaces determinados por el nivel de similitud entre ellos (de la Fuente, 2011). También se llama árbol de clasificación.

3.1.10. Selección de variables

La selección de las variables es un aspecto vital para determinar aquello que realmente contribuye a identificar los grupos requeridos con base en el propósito del estudio; se pueden utilizar para este tipo de análisis las variables cualitativas, ordinales o nominales, y las variables cuantitativas, discretas o continuas (de la Fuente, 2011).

3.1.11. Número óptimo de clústeres

El número óptimo de clústeres resultante del análisis de conglomerados es una decisión subjetiva que puede afectar la calidad en la conformación de cada agrupamiento. Esto se debe a que, para grandes volúmenes de datos, pocos clústeres pueden aumentar la heterogeneidad de cada conglomerado, mientras que muchos pueden complejizar la interpretación de los resultados. La decisión sobre el número apropiado de clústeres depende del criterio del investigador y sus requerimientos en términos de la separación entre conglomerados, por lo cual la decisión no siempre puede tomarse solo por los datos. Si bien la literatura sugiere que el número de conglomerados debe ser conocido o estimado a partir de los datos, fijar el número de clústeres por parte del investigador es una decisión legítima (Hennig, 2015). Existen varios criterios estadísticos para establecer el número adecuado de clúster, dentro de los que figuran: la técnica multivariada MANOVA; el criterio de Calinski y Harabasz; maximización de la traza $BW-1$; minimización del determinante W ; minimización de la traza W ; minimización de $K2^*$ y el determinante W .

3.1.12. Medida de la distancia de verosimilitud

La medida de la distancia de verosimilitud se realiza por medio de la distribución de probabilidad entre las variables. Para ello se parte del supuesto de que las variables del modelo de conglomeración son independientes, que cada variable continua tiene una distribución normal y que cada variable categórica tiene una distribución multinomial (de la Fuente, 2011). Para comprobar el cumplimiento de estos supuestos se pueden utilizar análisis como: correlaciones bivariadas para comprobar la independencia de las variables continuas; tablas de contingencia para comprobar la independencia de las variables categóricas; medias para comprobar la independencia entre una variable continua y otra variable categórica; exploración para verificar la normalidad de una va-

riable continua y la prueba de chi-cuadrado para comprobar si una variable categórica tiene una determinada distribución multinomial (de la Fuente, 2011).

3.1.13. Método de Ward

Es una técnica de agrupamiento que une los casos minimizando la varianza al interior de cada grupo al calcular la media de todas las variables en cada conglomerado y la distancia de cada caso con la media del conglomerado, posteriormente se suman las distancias entre todos los casos. Los conglomerados se agrupan al calcular la menor suma de las distancias dentro de cada conglomerado. La ventaja de este método es que genera grupos homogéneos y con tamaños similares (de la Fuente, 2011).

3.1.14. K-medias

El análisis de agrupamiento de k-medias es una técnica estadística multivariante especialmente utilizada en estudios económicos con el objetivo de reconocer los grupos que surgen de forma natural a partir de las observaciones. También se ha aplicado con éxito en el campo de las investigaciones sobre el desperdicio de alimentos (Amicarelli et al., 2021).

3.2. *Tendencia de la investigación sobre análisis de clúster*

3.2.1. Análisis del consumidor

Di Vita et al. (2021) utilizaron el análisis de clúster mediante el método de Ward para identificar la preferencia del consumidor respecto al aceite de oliva extra virgen en Italia, para lo cual utilizaron tres tipos de atributos: locales, orgánicos e indicación geográfica protegida. Encontraron cuatro grupos con características sociodemográficas diferentes.

Amicarelli et al. (2021) recurrieron a la técnica de agrupamiento de k-medias para analizar los tipos de desperdicio de alimentos en Italia con

base en las actitudes, percepciones y patrones de comportamiento de los consumidores. Encontraron tres grupos: los que declaran un alto nivel de conciencia sobre el desperdicio de alimentos pero desperdician mucho; los que tienen una ligera noción de sostenibilidad y desperdician pequeñas cantidades de alimentos y los que conocen el tema y lo ponen en práctica.

Chong et al. (2020) utilizaron el análisis de clúster para estudiar el gusto de los consumidores por la carne de res. Encontraron cuatro conglomerados de consumidores: fáciles de complacer, amantes de la carne de toro, amantes de la carne tierna y exigentes. Se puede concluir que esta herramienta permite establecer las preferencias del consumidor que pueden orientar estrategias de marketing.

Tleis et al. (2017) estudiaron los perfiles de consumidores de alimentos orgánicos en el Líbano con base en variables de estilo de vida y actitud, para ello utilizaron el análisis de componentes principales y de conglomerados por la técnica k-medias. Hallaron cuatro grupos en función de las características psicográficas y la disposición de los consumidores a pagar por los productos orgánicos, a saber: localistas, conscientes de la salud, racionales e irregulares. Se puede afirmar entonces que esta herramienta permite la orientación de estrategias de marketing adecuadas para cada segmento de mercado.

3.2.2. Análisis ambiental

En Ghana, Gyimah et al. (2021) evaluaron la calidad del agua de un río utilizando clustering jerárquico por el método de Ward y distancia euclidiana como parámetro de similitud entre clústeres. Analizaron parámetros fisicoquímicos y obtuvieron tres conglomerados de acuerdo con sus características de calidad del agua: bastante buena, buena y mala. Identificaron para cada grupo las fuentes de contaminación.

Tkachev I et al. (2020) utilizaron el análisis de conglomerados para llevar a cabo un modelado matemático del desarrollo de la industria agraria con base en las inversiones. Hallaron tres grupos homogéneos por las características de distribución de sus inversiones.

Wang et al. (2020) evaluaron el consumo de energía en China aplicando análisis de datos espaciales con agrupación jerárquica espacial. Identificaron tres grupos de consumo, a saber: las partes noreste, norte y este de China.

Abdelkawy et al. (2020) evaluaron características como: número de plantas/m², fertilidad de macollamiento, altura de planta (cm), longitud de espiga (cm), número de espiguillas/espiga, peso de grano (g) y rendimiento de grano (g/m²) de veinte genotipos de triticale en un diseño de celosía alfa y RCBD para comparar su eficiencia relativa en el diseño de celosía alfa y RCBD. Encontraron por medio del análisis de conglomerados tres grupos principales.

Al Saad & Hamdan (2020) analizaron el efecto de diferentes variables sobre la calidad del agua de plantas de tratamiento en Iraq. Utilizando análisis de componentes principales, análisis de factor y análisis de clúster encontraron tres conglomerados entre las estaciones muestreadas en términos de la calidad del agua utilizada para consumo e irrigación.

Con el propósito de identificar grupos con perfiles de resiliencia similares en comunidades en Noruega, Opach et al. (2020) realizaron un análisis de conglomerados para dividir los municipios con base en indicadores socioeconómicos y ambientales como medidas de la resiliencia. Encontraron seis grupos de municipios que se enmarcan en un clúster urbano y un clúster suburbano, lo cual les permitió concluir que la agrupación contribuye a la generación de redes entre los municipios con características

de resiliencia similares con el fin de compartir procesos de planificación para enfrentar peligros naturales compartidos por sus territorios.

Meschede et al. (2019) realizaron un análisis de conglomerados en 502 islas en Filipinas para clasificarlas según sus similitudes en características socioeconómicas y físicas y el potencial de recursos energéticos autóctonos. Hallaron cinco clústeres y concluyeron que los resultados del análisis de conglomerados contribuyen a la definición de políticas para el mejoramiento de sistemas de energía inteligente adecuados a las características homogéneas de las islas.

3.2.3. Salud

Čelakovská et al. (2020) emplearon el análisis de clúster para determinar componentes moleculares relacionados con pacientes con dermatitis atópica; encontraron diez clústeres con diferentes números de componentes moleculares asociados a familias de proteínas acordes con su estructura bioquímica.

Escalera-Antezana et al. (2020) utilizaron el análisis de conglomerados para estudiar una enfermedad emergente en Bolivia. Hallaron un conglomerado de cinco casos y llegaron a la conclusión que el análisis de clúster permitió identificar los factores socioambientales y ocupacionales que pueden contribuir al riesgo de contagio y la transmisión zoonótica a los humanos.

3.2.4. Marketing

Chayathatto et al. (2020) utilizaron el análisis de conglomerados para estudiar la propuesta de valor de los grupos de empresas de la comunidad agrícola en Tailandia. Encontraron tres grupos distinguibles de empresas: empresas en desarrollo, empresas avanzadas y empresas

principiantes de marketing digital, lo que les permitió concluir que esta clasificación puede ayudar a las empresas y ejecutivos en la planificación, ejecución o evaluación de este tipo de empresas agrícolas comunitarias.

De la Hoz & Polo (2017) estudiaron las condiciones competitivas en las organizaciones empresariales de comercio exterior de Colombia; a través del análisis de conglomerados consideraron 16 factores clave en el potencial exportador y encontraron cuatro conglomerados homogéneos con base en el potencial competitivo. Concluyeron que el análisis de conglomerados y de las redes neuronales artificiales son útiles para discriminar niveles competitivos en el potencial exportador.

3.3. El análisis de clúster en el sector agropecuario

3.3.1. Producción animal

En Pakistán, Fatih et al. (2021) utilizaron el análisis de conglomerados jerárquicos basados en la distancia euclidiana enlace simple para predecir el peso de los camellos con base en su caracterización morfológica. Esto se hizo a partir de la determinación de los rasgos morfológicos que más influyen de forma positiva en la ganancia de peso. Concluyeron que esta herramienta puede ayudar a los criadores a reproducir camellos de élite.

Baron et al. (2020) también utilizaron el análisis de clúster jerárquico para clasificar granjas porcícolas y mercados en grupos acordes con sus métricas en el análisis de redes sociales en Argentina, con lo cual obtuvieron cuatro conglomerados: granjas pequeñas y de traspatio, granjas industriales, mercados y mercado periférico. Concluyeron que este análisis brinda información valiosa para determinar estrategias y costos más efectivos para realizar programas de control y vigilancia epidemiológica tanto para enfermedades emergentes como endémicas.

Konkina & Martynushkin (2020) evaluaron el nivel de sustitución de importaciones en Rusia utilizando el análisis de clúster con base en indicadores técnicos y económicos que caracterizan el desarrollo de la industria láctea. Hallaron que las regiones se dividieron en dos grupos: el primero con gran éxito en la organización de esta industria desde la perspectiva de la sustitución de importaciones y el segundo con poco éxito en este mismo aspecto. La conclusión a la que llegaron es que esta herramienta permite puntualizar las acciones requeridas para mejorar la sustentabilidad de la industria e intensificar los procesos de sustitución de importaciones a través del sistema de regulación estatal.

3.3.2. Producción agrícola

Islam et al. (2020) utilizaron el análisis de conglomerados para estimar la variabilidad genética a partir del rendimiento del grano en Tailandia. Encontraron tres grupos de genotipos con mejores rasgos diferentes y concluyeron que esta herramienta es útil para los programas de mejoramiento de variedades de arroz seco.

Hloušková & Lekešová (2020) utilizaron el análisis de conglomerados para examinar la estructura de las explotaciones agrícolas en la República Checa, para lo cual emplearon cinco dimensiones: producción, factores económicos, estabilidad financiera, factores ambientales y sociales; realizaron la clasificación de las granjas en dos conglomerados con características diferenciadas. Concluyeron que esta herramienta constituye un insumo para el diseño de políticas agrícolas para la identificación de áreas en las cuales debe incrementarse la sustentabilidad para mejorar el sector agrícola.

Torrado & Sili (2020) usaron el análisis de conglomerados para identificar los modelos de toma de decisiones en el sector agropecuario en Argentina. Para ello estudiaron las características productivas y socio-

familiares de los productores agropecuarios y los procesos de toma de decisiones en sus explotaciones agropecuarias en aspectos técnico-productivos y económico-financieros. Identificaron cuatro conglomerados con diferentes modelos de toma de decisiones: innovador en red, familiar-tradicional, familiar-empresarial y directivo-administrativo.

Abdulrahman & Hameed (2020) utilizaron el método jerárquico del análisis de conglomerados para estudiar la producción de cítricos en Iraq. Hallaron tres conglomerados en los cuales se agrupan las regiones productoras de cítricos.

Silva et al. (2020) analizaron cultivos hidropónicos y convencionales de Jambú con el fin de determinar su composición química. Mediante el análisis de clúster encontraron tres conglomerados bien definidos: dos de partes anatómicas y el otro del sistema de cultivo.

Por medio del análisis de clúster, Ullah et al (2020) construyeron regiones climáticas homogéneas en Pakistán orientadas a optimizar el uso del agua. Así, identificaron cinco conglomerados homogéneos relacionados con las características de cada estación muestreada. Concluyen que esta técnica contribuye a la gestión de los recursos hídricos, al monitoreo de las sequías y sus proyecciones y, en consecuencia, a la adaptación, mitigación y planificación base en las proyecciones climáticas.

Butler et al. (2020) estudiaron las relaciones minerales-nutrientes del suelo para comprender su fertilidad utilizando el análisis de conglomerados con el algoritmo de agrupación *fuzzy-c-means* con un conjunto de datos de difracción de rayos X en polvo (XRPD) de 935 suelos del África subsahariana. Encontraron nueve conglomerados mineralógicamente distintos.

Seaton et al. (2020) estudiaron el cambio de los suelos a partir de la identificación de parámetros clave de su salud y funcionamiento en el

Reino Unido; hallaron cuatro clases funcionales de suelos con base en el análisis de conglomerados y concluyeron que este insumo constituye una base importante para el monitoreo de los suelos y las políticas agroambientales.

Akbari et al. (2020) estudiaron la desertificación en Irán a partir de indicadores, utilizando análisis de factores y opiniones de expertos. A través del análisis factorial priorizaron los parámetros efectivos, 27 de ellos que justificaron el 78% de la varianza total y el análisis de clúster, mediante el cual priorizaron cinco conglomerados. Concluyeron que el análisis factorial y el análisis de clúster son herramientas efectivas para la priorización e identificación de criterios de degradación de la tierra.

Maione et al. (2019) utilizaron el análisis de clúster para establecer los perfiles de las familias de pequeños productores y agricultores familiares en Brasil con base en información demográfica, económica, agrícola y de inseguridad alimentaria. Emplearon el método *Silhouette* para la estimación del mejor número de grupos y el algoritmo de agrupamiento *Partitioning Around Medoids* para calcular los perfiles; así obtuvieron una buena separación de las familias en dos grupos.

Gavioli et al. (2019) evaluaron 20 algoritmos para el análisis de conglomerados: vinculación promedio, agrupación en bolsa, vinculación centroide, clara, vinculación completa, Diana, Fanny, FCM, cáscaras C difusas, aprendizaje competitivo duro, agrupación jerárquica híbrida, K-medias, Método de McQuitty, vinculación mediana, Neural Gas, partición alrededor de medoides, enlace único, K-medias esféricas, aprendizaje competitivo difuso no supervisado y método de Ward. Para ello tomaron datos de cultivos de soja y maíz en Brasil y encontraron dos grupos que corresponden a clases de rendimientos significativamente diferentes. Concluyeron que las divisiones se realizaron satisfactoriamente con 17 algoritmos;

adicionalmente, resaltan que el método de McQuitty y Fanny fueron las mejores opciones al producir las mayores reducciones en la varianza y generar clases con alta homogeneidad interna.

Marino & Alvino (2018) utilizaron el análisis de conglomerados para la detección de áreas de trigo homogéneas, para lo cual emplearon imágenes de sensores remotos multitemporales y rasgos agronómicos de los cultivos. Concluyeron que el análisis de clúster, utilizando UAS y datos reales del suelo, es una buena estrategia para identificar las áreas homogéneas de cultivo de trigo.

Rodríguez et al. (2018) utilizaron el análisis de clúster para el estudio de la sostenibilidad de agroempresas asociativas rurales en Colombia. Hallaron tres conglomerados y concluyeron que esta herramienta constituye un insumo importante para la formulación de políticas públicas de fortalecimiento de la asociatividad rural.

Poudel et al. (2017) estudiaron cincuenta genotipos de trigo en Nepal para identificar variedades tolerantes al estrés por calor. Utilizaron el análisis de clúster para el agrupamiento y encontraron tres grupos homogéneos. Así, concluyeron que esta herramienta tiene mucha utilidad para la identificación y el desarrollo de variedades tolerantes al calor y otras actividades de mejoramiento.

3.4. Modelo de toma de decisiones con enfoque de clúster

Con base en los referentes analizados y el estado del arte de la utilización del análisis del clúster, se propone la siguiente metodología para la toma de decisiones en el sector agropecuario, más concretamente en el agroecoturismo, apoyada en el *clustering* y orientada a la clasificación de usuarios por medio de la identificación de las relaciones entre sus preferencias.

Paso 1: selección de las variables clave

Este paso es fundamental para garantizar la calidad del análisis y, en especial, para alcanzar los resultados esperados, pues permite identificar las variables que son relevantes para el análisis y las que realmente contribuyen a identificar los grupos requeridos a partir de la definición clara del propósito del estudio (de la Fuente, 2011); por otra parte, omitir variables de interés puede conducir a que los conglomerados obtenidos no sean los más apropiados. Una estrategia recomendable es la realización previa de un análisis factorial exploratorio para disminuir el grupo inicial de variables a partir de la identificación de los factores comunes.

Otro aspecto a tener en cuenta es la estandarización de las variables. De esta manera se disminuye el efecto que pueden tener las diferencias en el escalamiento de las variables cuando se tienen, por ejemplo, algunas variables en años, otras en pesos y otras en número de personas. En tal caso el paso se debe realizar en el análisis jerárquico de clúster.

Paso 2: definición del método de análisis de clúster

La definición del método de análisis para la obtención de los conglomerados debe realizarse teniendo en cuenta que cada algoritmo utiliza una forma diferente de medir la distancia entre clústeres. Por ello, la decisión que se debe tomar en este paso depende del método que refleje de mejor manera el propósito del estudio (de la Fuente, 2011). Es importante para tomar la decisión tener en cuenta que existen métodos jerárquicos (asociativos y disociativos) y no jerárquicos; que los jerárquicos se pueden utilizar para clasificar individuos o variables y que los no jerárquicos solo se pueden utilizar para clasificar individuos.

A pesar de la variedad de métodos existentes, se recomienda la utilización del método de Ward pues, al ser una técnica que minimiza la varianza al interior de cada grupo, genera grupos homogéneos y con tamaños simi-

lares (de la Fuente, 2011). También se recomienda el método de K-Medias, ampliamente utilizado en estudios económicos (Amicarelli et al., 2021).

Paso 3: selección del número de conglomerados

Para seleccionar el número óptimo de conglomerados se debe recurrir al criterio del investigador sobre el número de clústeres apropiado. Se debe hacer de acuerdo con el propósito del estudio y partiendo de los resultados que arrojen los datos, pero teniendo en cuenta que es legítimo que quien investiga fije el número de clústeres (Hennig, 2015).

Paso 4: análisis de las características de cada conglomerado

En el último paso el investigador toma la decisión de analizar las variables de cada conglomerado de una manera que le permita presentar de forma clara las características principales de cada agrupamiento, es decir, aquello que lo diferencia de los demás y que, por tanto, le permite alcanzar el resultado de clasificar usuarios a partir de los propósitos que se tengan en el estudio.

3.5. Caso de aplicación de segmentación de usuarios

Para dar mayor claridad sobre el modelo de toma de decisiones propuesto se presenta a continuación su aplicación a un caso para la clasificación de turistas con base en sus demandas por recursos ambientales. Para ello se tomó una base de datos de 499 personas con variables demográficas y preferencias por actividades e infraestructura disponibles en un destino turístico. Estas variables son dicotómicas, con medición nominal, y politómicas, con medición ordinal en escala tipo Likert de 1 a 5, siendo 5 el mayor nivel de preferencia.

3.5.1. Selección de las variables clave

Con base en los referentes de la literatura se propusieron 33 variables de interés para el análisis de las demandas de los turistas por recursos am-

bientales. Para reducir esa cantidad inicial de variables se realizó un análisis factorial exploratorio, el cual permitió definir las variables que tenían mayor impacto. Al comparar la nacionalidad de los individuos para las variables seleccionadas se encontró diferencia significativa ($p < 0,05$) entre los turistas nacionales y los extranjeros, con mayor nivel de preferencia por parte de los extranjeros (Tabla 1); la única variable en la que no se encontró diferencia significativa fue la percepción sobre las artesanías.

Tabla 1. Análisis descriptivo por tema y procedencia

Variable	Extranjero	Nacional
	Media \pm SD	Media \pm SD
Áreas verdes	3,87 \pm 0,99 a	3,49 \pm 1,29 b
Cuerpos agua	3,57 \pm 1,08 a	3,04 \pm 1,40 b
Avistamiento de aves	3,47 \pm 1,19 a	2,63 \pm 1,30 b
Avistamiento de flora	3,64 \pm 1,19 a	2,72 \pm 1,33 b
Artesanías	2,76 \pm 1,13 a	2,71 \pm 1,34 a
Educación ambiental	2,62 \pm 1,38 a	2,35 \pm 1,40 b
Rutas cafeteras	3,34 \pm 1,25 a	2,90 \pm 1,32 b
Arquitectura	3,41 \pm 1,20 a	3,07 \pm 1,31 b
Caminos de campo	3,71 \pm 1,11 a	3,19 \pm 1,27 b
Cultura rural	3,58 \pm 1,17 a	3,05 \pm 1,41 b
Rutas de fincas	3,43 \pm 1,19 a	2,73 \pm 1,35 b

Letras diferentes indican diferencia estadística significativa ($p < 0,05$)

Fuente: elaboración propia

Para perfeccionar la selección de las variables clave se realizó un análisis de correlación de tipo bidimensional mediante la técnica no paramétrica de Spearman, lo cual permitió determinar las mayores relaciones directamente proporcionales. Para las personas provenientes del exterior se encontró: existencia de zonas verdes y avistamiento de aves; cuerpos de agua con presencia de aves y flora y visita a fincas cafeteras con muestras culturales de tipo rural. Para las personas nacionales la mayor

correlación se dio entre: áreas verdes con presencia de flora, fauna, educación ambiental, arquitectura y cultura rural. Adicionalmente, se asoció a las rutas cafeteras con otras variables como la caminata por el campo, por las rutas de las fincas y la presencia de la cultura rural (Tabla 2).

Tabla 2. Análisis de correlación por el método de Spearman

	Extranjeros										
	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11
V1	1,0	0,61	0,42	0,41	0,18	0,28	0,24	0,21	0,19	0,18	0,23
V2		1,0	0,52	0,43	-0,01	0,11	0,02	0,08	0,18	-0,01	0,06
V3			1,0	0,51	0,00	0,09	-0,01	0,03	0,18	0,08	0,14
V4				1,0	0,14	0,13	0,11	0,11	0,23	0,12	0,11
V5					1,0	0,48	0,32	0,39	0,16	0,19	0,31
V6						1,0	0,36	0,31	0,09	0,15	0,32
V7							1,0	0,38	0,17	0,30	0,50
V8								1,0	0,40	0,42	0,32
V9									1,0	0,36	0,34
V10										1,0	0,53
V11											1,0
	Nacionales										
	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11
V1	1,0	0,53	0,47	0,48	0,26	0,44	0,41	0,56	0,28	0,47	0,36
V2		1,0	0,46	0,42	0,13	0,35	0,28	0,34	0,40	0,33	0,30
V3			1,0	0,64	0,26	0,41	0,29	0,38	0,27	0,32	0,31
V4				1,0	0,26	0,48	0,31	0,38	0,25	0,44	0,27
V5					1,0	0,50	0,43	0,41	0,26	0,35	0,48
V6						1,0	0,49	0,43	0,29	0,44	0,45
V7							1,0	0,65	0,35	0,50	0,54
V8								1,0	0,50	0,53	0,45
V9									1,0	0,35	0,33
V10										1,0	0,61
V11											1,0

V1: Áreas verdes; V2: Cuerpos agua; V3: Avistamiento de aves; V4: Avistamiento flora; V5: Artesanías; V6: Educación ambiental; V7: Rutas cafeteras; V8: Arquitectura; V9: Caminatas campo; V10: Cultura rural; V11: Rutas fincas

Fuente: elaboración propia

Además de lo anterior, se efectuó el análisis multivariado de la varianza con la técnica MANOVA con contraste canónico; se determinó por la técnica de máxima verosimilitud el valor propio idóneo para efectuar la comparación entre nacionales y extranjeros, al igual que entre géneros, utilizando el paquete estadísticos SAS University. Al contrastar las preferencias entre nacionales y extranjeros se detectó diferencia altamente significativa (Tabla 3), a saber, que los extranjeros tienen mayor preferencia por los recursos agroecoturísticos. Igualmente, se identificó una divergencia entre géneros: las mujeres tienen mayor preferencia por este tipo de recursos.

Tabla 3. Análisis multivariado de la varianza MANOVA

		Comparación entre nacionales y extranjeros
MANOVA		
Wilks' Lambda		<0,0001
Pillai's Trace		<0,0001
Hotelling-Lawley Trace		<0,0001
Roy's Greatest Root		<0,0001
Contraste	Donde difieren	
Tukey	Áreas verdes	Mayor nota extranjeros
Tukey	Cuerpos agua	
Tukey	Avistamiento de Flora	
Tukey	Educación Ambiental	
Tukey	Rutas Cafeteras	
Tukey	Arquitectura	
Tukey	Caminata de campo	
Tukey	Cultura rural	
Tukey	Rutas fincas	
Tukey	Avistamiento aves	

Fuente: elaboración propia

3.5.2. Definición del método de análisis de clúster

El análisis de clúster se llevó a cabo empleando el algoritmo de clúster jerárquico denominado método de Ward (Hair et al., 2014). En este, los conglomerados se forman mediante la fusión de dos elementos con la característica de que la pérdida de información se minimiza. Es importante anotar que se tiene en cuenta la suma de cuadrados del error. Como medida de similitud se aplicó la distancia euclídea al cuadrado.

3.5.3. Selección del número de conglomerados

El análisis de conglomerados permitió identificar que existe variabilidad entre el conjunto de usuarios extranjeros respecto a los nacionales, lo cual ratifica los resultados asociados con el análisis multivariado de la varianza. Debido a que se presentaron diferencias multivariadas entre nacionales y extranjeros se realizó el análisis de clúster de manera independiente. El análisis de clúster jerárquico sugirió un grupo óptimo de 3 conglomerados, que a juicio de los autores son apropiados para los propósitos del estudio (Figura 1).

3.5.4. Análisis de las características de cada conglomerado

Las personas de otras nacionalidades prefieren en primera instancia las rutas cafeteras en las que se destaca su arquitectura y hay presencia de zonas verdes. En segundo lugar, prefieren la presencia de cuerpos de agua y avistamiento de flora y fauna; luego las artesanías y la educación ambiental y, con un peso inverso sobre el factor, las caminatas de campo y la cultura rural. Para las personas nacionales el clúster uno básicamente lo definen casi todas las variables (Tabla 4).

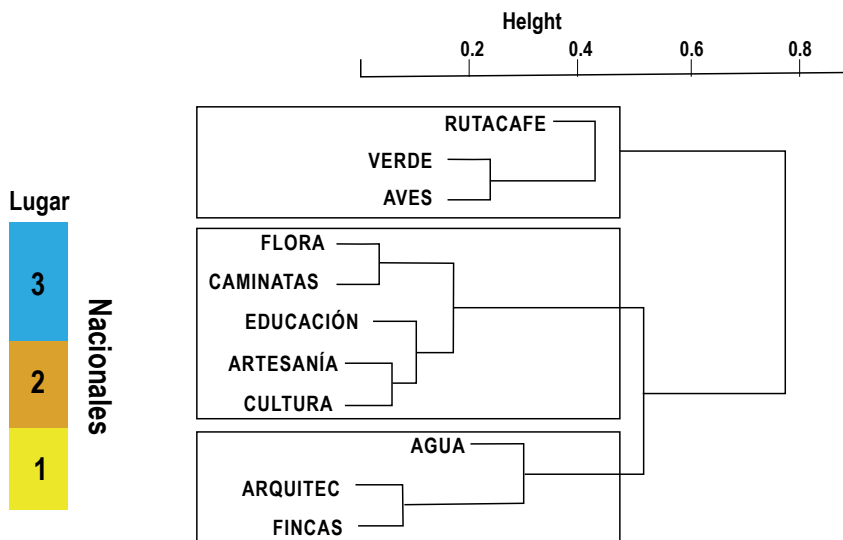
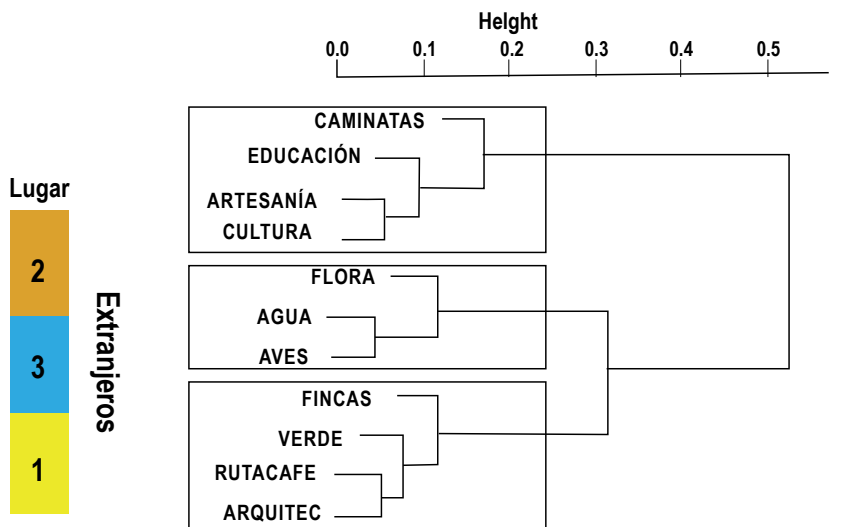


Figura 1. Análisis de clúster por tema

Fuente: elaboración propia

Tabla 4. Preferencias de usuarios por clúster

Variable	Extranjeros			Nacionales		
	Clúster 1	Clúster 2	Clúster 3	Clúster 1	Clúster 2	Clúster 3
Ambiental						
Áreas verdes	0,645	0,431	0,169	0,716	0,285	0,019
Cuerpos agua	0,454	0,697	0,083	0,611	0,479	0,290
Avista. aves	0,455	0,653	-0,026	0,657	0,476	-0,219
Avist. flora	0,498	0,536	0,027	0,676	0,449	-0,305
Artesanías	0,509	-0,371	0,477	0,576	-0,456	-0,357
Ed. ambiental	0,542	-0,255	0,593	0,737	-0,066	-0,289
Rut. cafeteras	0,592	-0,40.1	0,114	0,722	-0,367	0,075
Arquitectura	0,627	-0,327	-0,117	0,771	-0,176	0,241
Camin. campo	0,523	0,022	-0,533	0,548	-0,041	0,638
Cultura rural	0,586	-0,307	-0,468	0,733	-0,174	0,024
Rutas fincas	0,675	-0,328	-0,228	0,711	-0,353	-0,041
Acumulado	0.313	0,501	0,614	0,465	0,580	0,664

Fuente: elaboración propia

4. Conclusiones

El estudio de los referentes de la literatura sobre análisis de conglomerados permitió proponer una metodología para la toma de decisiones en el sector agropecuario, apoyada en el agrupamiento, utilizando el método de clúster. Esta metodología fue validada para un caso de toma de decisiones sobre servicios agroecoturísticos que se deben ofrecer a los visitantes de un municipio a partir del reconocimiento de sus preferencias, lo cual permitió su clasificación en segmentos de mercado por medio de la identificación de las relaciones entre sus preferencias.

El análisis de clúster es una herramienta útil para la toma de decisiones en procesos de innovación en el sector agroalimentario, por cuanto permite diseñar productos y servicios diferenciados que se ajustan mejor a las necesidades de los usuarios potenciales como en el caso presenta-

do, consumidores de bienes y servicios generados por los productores agropecuarios. En este sentido, la metodología propuesta se constituye en un insumo que puede ser utilizado, por ejemplo, para planificar la política de innovación agropecuaria con base en la segmentación de mercados y la diversificación de productos y servicios de extensión agropecuaria, teniendo en cuenta las necesidades de los productores.

Referencias bibliográficas

- Abdelkawy, R., Turbayev, A. & Soloviev, A. (2020). Evaluation of Relative Efficiency of Alpha Lattice Design and Cluster Analysis of Twenty Spring Triticale Field Experiments. *Indian Journal of Agricultural Research*, 54, 301–307. <https://arccjournals.com/journal/indian-journal-of-agricultural-research/A-466>
- Abdulrahman, H. & Hameed, L. (2020). The cluster analysis of most important citrus trees in some governorates of Iraq for the year 2019. *Periodicals of Engineering and Natural Sciences*, 8(4), 1999–2006. <https://doi.org/10.21533/PEN.V8I4.1676.G683>
- Akbari, M., Koushki, F.F., Memarian, H., Azamirad, M., Mohammad & Noughani, A. (2020). Prioritizing effective indicators of desertification hazard using factor-cluster analysis, in arid regions of Iran. *Arabian Journal of Geosciences*, 13(319), 1–17. <https://doi.org/10.1007/s12517-020-05296-9>
- Al Saad, Z. A. A. & Hamdan, A. N. A. (2020). Evaluation of water treatment plants quality in Basrah Province, by factor and cluster analysis. *Journal of Water and Land Development*, 46(7–9), 10–19. <https://doi.org/10.24425/jwld.2020.134097>
- Amicarelli, V., Tricase, C., Spada, A. & Bux, C. (2021). *Households' Food Waste Behavior at Local Scale: A Cluster Analysis after the COVID-19 Lockdown*. <https://doi.org/10.3390/su13063283>
- Arista-Jalife, A., Calderón-Auza, G., Fierro-Radilla, A. & Nakano, M. (2017). Clasificación de imágenes urbanas aéreas: Comparación entre descriptores de bajo nivel y aprendizaje profundo. *Información Tecnológica*, 28(3), 209–224. <https://doi.org/10.4067/S0718-07642017000300021>

- Baron, J. N., Aznar, M. N., Monterubbianesi, M. & Martínez-López, B. (2020). Application of network analysis and cluster analysis for better prevention and control of swine diseases in Argentina. *PLoS ONE*, 15(6), 1–26. <https://doi.org/10.1371/journal.pone.0234489>
- Butler, B. M., Palarea-Albaladejo, J., Shepherd, K. D., Nyambura, K. M., Towett, E. K., Sila, A. M. & Hillier, S. (2020). Mineral–nutrient relationships in African soils assessed using cluster analysis of X-ray powder diffraction patterns and compositional methods. *Geoderma*, 375, 114474. <https://doi.org/10.1016/j.geoderma.2020.114474>
- Carpaneto, E., Chicco, G., Napoli, R. & Scutariu, M. (2006). Electricity customer classification using frequency-domain load pattern data. *International Journal of Electrical Power and Energy Systems*, 28(1), 13–20. <https://doi.org/10.1016/j.ijepes.2005.08.017>
- Čelakovská, J., Bukač, J., Vaňková, R., Cermakova, E., Krcmova, I., Krejsek, J. & Andrýs, C. (2020). Cluster analysis of molecular components in 100 patients suffering from atopic dermatitis according to the ISAC Multiplex testing. <https://doi.org/10.1080/09540105.2020.1776224>
- Chayathatto, Mada Na Krom Woradech, A. & Abdul, R. (2020). Using Cluster Analysis to Understand the Value Proposition in Digital Business to Develop Digital Marketing for Agricultural Community Enterprise in Thailand. *Solid State Technology*, 63(6). <http://solidstatetechnology.us/index.php/JSST/article/view/3923>
- Chong, F., O'Sullivan, M., Kerry, J., Moloney, A., Methven, L., Gordon, A., Hagan, T. & Farmer, L. (2020). Understanding consumer liking of beef using hierarchical cluster analysis and external preference mapping. <https://onlinelibrary.wiley.com/doi/abs/10.1002/jsfa.10032>
- de la Fuente, S. (2011). *Análisis de conglomerados*. Universidad Autónoma de Madrid. http://www.estadistica.net/Master-Econometria/Analisis_Cluster.pdf
- de la Hoz, E., Mendoza, A. & Ojeda, H. (2017). Clasificación de perfiles de lectores de un periódico digital. *Revista U.D.C.A Actualidad & Divulgación Científica*, 20(2), 469–478. <https://biblat.unam.mx/es/revista/revista-udca-actualidad-divulgacion-cientifica/articulo/clasificacion-de-perfiles-de-lectores-de-un-periodico-digital>

- de la Hoz, E. & Polo, L. L. (2017). Aplicación de Técnicas de Análisis de Conglomerados y Redes Neuronales Artificiales en la Evaluación del Potencial Exportador de una Empresa. *Informacion Tecnologica*, 28(4), 67–74. <https://doi.org/10.4067/S0718-07642017000400009>
- di Vita, G., Zanchini, R., Falcone, G., D'Amico, M., Brun, F. & Gulisano, G. (2021). Local, organic or protected? Detecting the role of different quality signals among Italian olive oil consumers through a hierarchical cluster analysis. *Journal of Cleaner Production*, 290, 125795. <https://doi.org/10.1016/j.jclepro.2021.125795>
- Escalera-Antezana, J. P., Rodriguez-Villena, O. J., Arancibia-Alba, A. W., Alvarado-Arnez, L. E., Bonilla-Aldana, D. K. & Rodríguez-Morales, A. J. (2020). Clinical features of fatal cases of Chapare virus hemorrhagic fever originating from rural La Paz, Bolivia, 2019: A cluster analysis. *Travel Medicine and Infectious Disease*, 36, 101589. <https://doi.org/10.1016/j.tmaid.2020.101589>
- Fatih, A., Celik, S., Eyduran, E., Tirink, C., Tariq, M. M., Shahzad Sheikh, I., Faraz, A. & Waheed, A. (2021). Use of MARS algorithm for predicting mature weight of different camel (*Camelus dromedarius*) breeds reared in Pakistan and morphological characterization via cluster analysis. *Tropical Animal Health and Production*, 53(191), 1–14. <https://doi.org/10.1007/s11250-021-02633-2>
- Fontalvo-Herrera, T. J., Delahoz, E. J. & Mendoza-Mendoza, A. A. (2018). Application of data mining for the classification of university programs of industrial engineering accredited in high quality in Colombia. *Informacion Tecnologica*, 29(3), 89–96. <https://doi.org/10.4067/S0718-07642018000300089>
- Gavioli, A., de Souza, E. G., Bazzi, C. L., Schenatto, K., & Betzek, N. M. (2019). Identification of management zones in precision agriculture: An evaluation of alternative cluster analysis methods. *Biosystems Engineering*, 181, 86–102. <https://doi.org/10.1016/j.biosystemseng.2019.02.019>
- Gyimah, R. A. A., Gyamfi, C., Anornu, G. K., Karikari, A. Y. & Tsyawo, F. W. (2021). Multivariate statistical analysis of water quality of the Densu River, Ghana. In *International Journal of River Basin Management* (Vol. 19, Issue 2, pp. 189–199). <https://doi.org/10.1080/15715124.2020.1803337>

- Hennig, C. (2015). Clustering strategy and method selection. *Handbook of Cluster Analysis*, 703–730. <https://doi.org/10.1201/b19706>
- Hloušková, Z. & Lekešová, M. (2020). Farm outcomes based on cluster analysis of compound farm evaluation. *Agricultural Economics (Czech Republic)*, 66(10), 435–443. <https://doi.org/10.17221/273/2020-AGRICECON>
- Islam, S., Anothai, J., Nualsri, C. & Soonsuwon, W. (2020). Genetic Variability and Cluster Analysis for Phenological Traits of Thai Indigenous Upland Rice (*Oryza sativa* L.). *Indian Journal of Agricultural Research*, 54(2), 211–216.
- Knijnenburg, B. P., Willemsen, M.C., Willemsen, C., Gantner, Z., Soncu, H., Newell, C., Knijnenburg, B. P., Willemsen, · M C, Gantner, Z., Soncu, H. & Newell, C. (2012). *Explaining the user experience of recommender systems*. 22, 441–504. <https://doi.org/10.1007/s11257-011-9118-4>
- Konkina, V. & Martynushkin, A. (2020). Analysis of Import Substitution Processes in the Milk and Dairy Products Market Using Cluster Analysis. *International Transaction Journal of Engineering, Management, & Applied Sciences & Technologies*, 11(10), 1–10. <https://doi.org/10.14456/ITJEMAST.2020.195>
- Krippendorff, K. (2004). Reliability in content analysis: Some common misconceptions and recommendations. *Human Communication Research*, 30, 411–433.
- Lamnek, S. (2005). *Qualitative social research*. Holz.
- Maione, C., Nelson, D. R. & Barbosa, R. M. (2019). Research on social data by means of cluster analysis. *Applied Computing and Informatics*, 15(2), 153–162. <https://doi.org/10.1016/j.aci.2018.02.003>
- Marino, S. & Alvino, A. (2018). Detection of homogeneous wheat areas using multi-temporal UAS images and ground truth data analyzed by cluster analysis. *European Journal of Remote Sensing*, 51(1), 266–275. <https://doi.org/10.1080/22797254.2017.1422280>
- Meschede, H., Esparcia, E. A., Holzapfel, P., Bertheau, P., Ang, R. C., Blanco, A. C. & Ocon, J. D. (2019). On the transferability of smart energy systems on off-grid islands using cluster analysis – A case study for the Philippine archipelago. *Applied Energy*, 251, 113290. <https://doi.org/10.1016/j.apenergy.2019.05.093>

- Molina, O. M. (2007). La segmentación de la demanda turística española. *Metodología de Encuestas*, 7, 57–92. <http://casus.usal.es/pkp/index.php/MdE/article/view/981>
- Opach, T., Scherzer, S., Lujala, P. & Rød, J. K. (2020). Seeking commonalities of community resilience to natural hazards: A cluster analysis approach. *Norsk Geografisk Tidsskrift-Norwegian Journal of Geography*, 74(3), 181–199. <https://doi.org/10.1080/00291951.2020.1753236>
- Poudel, A., Thapa, D. B. & Sapkota, M. (2017). Cluster Analysis of Wheat (*Triticum aestivum* L.) Genotypes Based Upon Response to Terminal Heat Stress. *International Journal of Applied Sciences and Biotechnology*, 5(2), 188–193. <https://doi.org/10.3126/ijasbt.v5i2.17614>
- Rodríguez, H., Ramírez, C. J. & Restrepo, L. F. (2018). Factores Determinantes de la Sostenibilidad de las Agroempresas Asociativas Rurales. *Revista Economía e Sociología Rural*, 56(1), 107–122. <http://www.scielo.br/pdf/resr/v56n1/1806-9479-resr-56-01-107.pdf>
- Seaton, F. M., Barrett, G., Burden, A., Creer, S., Fitos, E., Garbutt, A., Griffiths, R. I., Henrys, P., Jones, D. L., Keenan, P., Keith, A., Lebron, I., Maskell, L., Pereira, M.G., Reinsch, S., Smart, S. M., Williams, B., Emmett, B. A. & Robinson, D. A. (2020). *Soil health cluster analysis based on national monitoring of soil indicators*. <https://doi.org/10.1111/ejss.12958>
- Silva, L., Aceval, N., Lemos, L., Faqueti, L., Sandjo, L., Salles, C., Weber, M., Barcelos-Oliveira, J. & Dias, R. (2020). Phytochemical profile of different anatomical parts of jambu (*Acmella oleracea* (L.) R.K. Jansen): A comparison between hydroponic and conventional cultivation using PCA and cluster analysis. *Food Chemistry*, 332, 127393. <https://doi.org/10.1016/j.foodchem.2020.127393>
- Tkachev I, S., Berdnova V, E., Rubtsova, S. N., Pakhomova, T. V., Lazhauninkas, J. V. & Sleptsova, L. A. (2020). Economic Mathematical Modeling of Agrarian Industry Development by Cluster Analysis. *Turismo-Estudos E Praticas*, 1, 1–13.
- Tleis, M., Callieris, R. & Roma, R. (2017). Segmenting the organic food market in Lebanon: an application of k-means cluster analysis. *British Food Journal*, 119(7), 1423–1441. <https://doi.org/10.1108/BFJ-08-2016-0354>

- Torrado, R. & Sili, M. (2020). Toma de decisiones y gestión productiva en el sector agropecuario del Noreste de La Pampa (Argentina). *Revista de Economía e Sociología Rural*, 58(2), 1–19. <https://doi.org/10.1590/1806-9479.2020.198357>
- Wang, S., Liu, H., Pu, H. & Yang, H. (2020). Spatial disparity and hierarchical cluster analysis of final energy consumption in China. *Energy*, 197, 117195. <https://doi.org/10.1016/j.energy.2020.117195>
- Ye, Q., Zhang, Z. & Law, R. (2009). Sentiment classification of online reviews to travel destinations by supervised machine learning approaches. *Expert Systems with Applications*, 36(3 PART 2), 6527–6535. <https://doi.org/10.1016/j.eswa.2008.07.035>