

Segmentación de mercado: Machine Learning en marketing en contextos de covid-19

PEDRO PABLO CHAMBI CONDORI¹

RECIBIDO: 15/09/2022 ACEPTADO: 14/02/2023 PUBLICADO: 31/07/2023

RESUMEN

La crisis sanitaria covid-19 ha provocado cambios jamás vistos en el comportamiento de los consumidores, quienes compran de manera diferente y por medios también diferentes. Los consumidores están mirando y valorando los productos a través de dispositivos electrónicos, configurando movimientos en segmentos de consumidores. El objetivo del presente estudio fue aplicar el modelo de clustering con recursos de Machine Learning en el análisis de conglomerados como recurso para la segmentación de consumidores, como un componente importante para la gestión del marketing empresarial. Para dicho propósito, se suministró un cuestionario de 6 preguntas a 506 personas de entre 18 y 65 años para recoger sus percepciones sobre el hecho de salir a comprar. Con los datos recogidos se organizó una dataset para procesarlo en el software RapidMiner Studio 9.10. Como resultado, se obtuvo la cantidad óptima de conglomerados y sus componentes a partir del indicador de performance procurado por Machine Learning.

Palabras clave: investigación de mercados; segmentación; inteligencia artificial; covid-19.

INTRODUCCIÓN

El *clustering* es un conjunto de técnicas estadísticas descriptivas no inferencial, que es del tipo de aprendizaje automático no supervisado en experiencias de *machine learning*, cuya aplicación está orientada a agrupar objetos similares en un mismo clúster, por lo que forman conglomerados con alta afinidad interna y diferencia externa. En marketing empresarial, la segmentación de mercado proporciona información acerca de la composición del mercado al diferenciar grupos con características y necesidades similares. Por otro lado, para una empresa, la segmentación de clientes es una herramienta fundamental que permite a la empresa identificar características comunes y diferenciadas dentro de la base de datos de clientes, las cuales son de mucha utilidad para la gestión estratégica de clientes.

En relación con el comportamiento de los consumidores en el contexto de la pandemia, Casco (2020) analizó el impacto del covid-19 en el comportamiento del consumidor y constató que estos recurren cada vez más a la tecnología para adquirir bienes y servicios, así como a servicios de reparto para la entrega de productos. Además, Mehta et al. (2020) estudiaron el comportamiento de los consumidores en el contexto del covid-19 y encontraron diferencias entre lo que ocurría en tiempos normales con lo que venía ocurriendo en tiempos de crisis, así como necesidades cambiantes provocadas por factores culturales y la edad de los clientes, que repercuten en nuevos modelos de comportamiento del consumidor. Eger et al. (2021) estudiaron las tendencias y el impacto del covid-19 en el comportamiento del consumidor, documentaron patrones de comportamiento en el transcurso de la segunda ola de la pandemia en el contexto de la República Checa e identificaron similitudes y diferencias intergeneracionales.

El brote de coronavirus ha sido una tragedia humana con impactos negativos en la economía global que ha alterado los conceptos de marketing, el contexto mismo de las tácticas y las estrategias del marketing (Kumar et al., 2020). Esta crisis sin precedentes obligó a las personas a aislarse, lo que alteró todos

¹ Doctor. Actualmente, es docente asociado de la Escuela de Ingeniería Comercial de la Universidad Nacional Jorge Basadre Grohmann de Tacna (Tacna, Perú).
Orcid: <https://orcid.org/0000-0002-8628-6825>
E-mail: pchambic@unjbg.edu.pe

los aspectos de sus actividades regulares (Kabir y Madria, 2021). La pandemia ha causado serios estragos, ante esta situación, muchos investigadores de diferentes campos del conocimiento unieron sus esfuerzos para proporcionar una amplia gama de soluciones, entre ellos el impulso del marketing digital (Rodríguez-Rodríguez et al., 2021). Yildirim et al. (2021) estudiaron el impacto negativo del covid-19 y demostraron que la vulnerabilidad, el riesgo percibido y el miedo pueden aumentar significativamente la presencia de comportamientos preventivos en los consumidores, lo que evidencia la variación dinámica en la agrupación de consumidores, estadios que deben ser estudiados mediante la segmentación de mercado y de clientes para la gestión estratégica de clientes.

El nivel de impacto de la aplicación de tecnologías de 4° generación en la gestión de mercados, en la configuración de calidad, diseño y desarrollo de productos y comunicación con sus clientes es cada vez más significativo. Al respecto, Brei (2020) explica que el marketing tiene muchas aplicaciones potenciales para el aprendizaje automático. En otras palabras, el marketing puede transformarse en un trabajo científico más autónomo mediante el uso de datos y la adecuada formulación de cada aplicación. El marketing tiene dos principales paradigmas tradicionales: el marketing de destino y el marketing de relaciones. Cada uno de los numerosos aspectos de los dos paradigmas de marketing se puede formular en un problema de aprendizaje automático. Es decir, para cada problema, se puede construir un modelo de aprendizaje automático y se pueden establecer parámetros del modelo. Las expresiones de inteligencia artificial han merecido estudios porque las máquinas con capacidades de aprendizaje profundo pueden llevar al marketing digital a niveles superiores y marcar la diferencia (Miklosik et al., 2019). En estos últimos tiempos, el marketing digital está en constante evolución, introduciendo nuevas herramientas para el tratamiento de datos que describen los nuevos hábitos de consumo, en donde el éxito o el fracaso de la comunicación digital depende en gran medida en la calidad de su marketing de contenidos (Baltes, 2015).

El impacto generalizado de la inteligencia artificial (IA) y el machine learning (ML) en muchos segmentos de la sociedad se ha sentido con fuerza en el campo del marketing (Jarek y Mazurek, 2019). En efecto, el machine learning ofrece una serie de resultados beneficiosos, como la oportunidad de utilizar técnicas sólidas para la generalización de los logros obtenidos por la ciencia y la investigación científica. Por otro lado, Ullal et al. (2021) afirman

que el machine learning es una disciplina que combina la ciencia, la estadística y la codificación computacional para detectar patrones en grandes volúmenes de datos. A partir de este soporte de datos, es posible realizar estimaciones a futuro, como por ejemplo estimados de ventas esperadas, cuota de atención a segmentos de mercados, gestión de clientes bajo el concepto de CRM, que resultan de haber aplicado algoritmos matemáticos y estadísticos. La integración del *big data* con las técnicas de aprendizaje automático es el resultado de tres factores clave: la disponibilidad de algoritmos y datos, la mayor capacidad de procesamiento de las computadoras y el menor costo de almacenamiento de los dispositivos digitales. Sin embargo, la velocidad de los avances en ciencias de la computación e ingeniería de software plantean un desafío que cada vez cobra mayor importancia: identificar e implementar la combinación adecuada de plataformas de hardware y arquitecturas de software, que aseguren la continuidad del modelo integrado y minimicen su obsolescencia, mitigando una pérdida de la capacidad de procesamiento de la plataforma de gestión de datos.

El objetivo de la presente investigación fue desarrollar un estudio de caso de la aplicación de uno de los recursos de inteligencia artificial en la investigación de mercados, específicamente, la segmentación de clientes como herramienta para la gestión estratégica del marketing empresarial.

Se hizo uso del modelo de clustering para la aplicación de segmentación de mercado. El fundamento del análisis clúster consiste en agrupar objetos formando conglomerados, aplicando los algoritmos de clustering como el modelo *k-means*. Este es el más popular de los algoritmos, el cual agrupa datos numéricos y en cada clúster tiene un centro llamado la media. En el algoritmo se asume que el número de clústers *k* es un valor fijo.

El citado algoritmo *k-means*, según Mahendiran et al. (2012), consiste en el análisis de distancia euclidiana al cuadrado entre dos variables que se define como la suma de cuadrados de las diferencias de todas las coordenadas de dos puntos como se observa en la siguiente ecuación:

$$d(i, j) = \sqrt{\sum_k (X_{ik} - X_{jk})^2}$$

Donde X_{ik} y X_{jk} son los valores individuales de cada uno de los puntos y el índice de Davies-Bouldin, que calcula las distancias que existe entre los conglomerados, se utiliza para evaluar el rendimiento.

En su estudio *Estimating Marketing Component Effects: Double Machine Learning from Targeted Digital Promotions*, Ellickson et al. (2022, p. 2), en primer lugar, buscaron demostrar cómo se puede aprovechar el aprendizaje automático en un marco unificado para proporcionar estimaciones causales del efecto de componentes separados en una intervención de marketing de alta dimensión, utilizando únicamente datos observacionales. En segundo lugar, en el estudio de referencia, la metodología se aplicó a los datos de 34 promociones de correo electrónico distintas enviadas por un minorista a más de 1.3 millones de personas en su lista de contactos para resaltar los hallazgos cuantitativos, los cuales son relevantes tanto para la literatura de marketing como para la práctica de gerencia. Además, Dzyabura y Hauser (2011) afirman que las reglas de decisión heurísticas predicen las decisiones de validación mejor que las reglas compensatorias.

Por su parte, Jordan y Mitchell (2015, p. 255) aseveran que el machine learning contribuye a que los fabricantes de computadoras construyan equipos inteligentes que incorporen la experiencia para potenciar sus aplicaciones. Este campo técnico ha experimentado un avance significativo en estos últimos años, que se ha desarrollado en la confluencia de la ciencia informática y los métodos cuantitativos, y en el centro de la inteligencia artificial y el procesamiento de datos. El acelerado avance del machine learning (aprendizaje automático) ha sido impulsado tanto por la rápida incursión de algoritmos innovadores, lenguajes de programación, mayor capacidad en los procesadores y la expansión de las teorías de aprendizaje, como el crecimiento continuo en la accesibilidad a datos en línea, repositorios de datos y recursos de computación disponibles de bajo costo. La adopción de métodos de machine learning con un uso intensivo de datos es evidente en todas las ciencias, en la gestión de negocios, en el comercio, en las finanzas y economía, en la gestión de la salud, en la educación, en vigilancia mediante la visión artificial y en la ingeniería, lo que ha contribuido a la mejora de la calidad de las decisiones basada en análisis de datos y proporciona evidencias para una toma de decisiones de calidad.

Al decir de Jain y Kain (2018), el aprendizaje automático es la ciencia que permite que las computadoras trabajen de manera automatizada. En los últimos diez años, el aprendizaje automático nos ha proporcionado automóviles autónomos, aplicaciones de reconocimiento de voz, búsqueda web efectiva y una comprensión considerablemente

mejorada del genoma humano. El estudio se centró en el método para predecir el comportamiento de mercado de valores utilizando algoritmos de aprendizaje automático: regresión lineal, bosque aleatorio y perceptrón multicapa.

Abakouy et al. (2019) realizaron un estudio comparativo sobre los métodos de aprendizaje automático más populares aplicados al desafiante problema de la personalización del marketing por correo electrónico. Las líneas de asunto y remitente tienen una gran influencia en las tasas de clics de los correos electrónicos, ya que los clientes a menudo abren y hacen clic en correos electrónicos en función de ellos. Según Kaličanin et al. (2019, p. 472), la inteligencia artificial (IA) impacta numerosos aspectos de la vida en forma de dispositivos y aplicaciones inteligentes, diseñados para comprender el comportamiento, las necesidades y las preferencias del consumidor con el fin de brindar experiencias personalizadas. La IA ha sido uno de los principales impulsores de la innovación en marketing. Los profesionales del marketing ya están aprovechando las ventajas de la IA para obtener información valiosa sobre los clientes, la competencia y los mercados. La IA automatiza tareas, reduce costos y mejora los flujos de trabajo. En su artículo académico, los autores examinan las aplicaciones actuales y potenciales de la IA dentro del marketing y proporcionan una visión general completa de la investigación académica existente.

Sobre el enfoque inteligente del marketing, Mahajan et al. (2017, p. 21929) sostienen que la IA es un sistema de soporte de gestión de marketing de vanguardia para tratar el conocimiento mediante el aprendizaje automático y otras técnicas de computación blanda. La gama de aplicaciones potenciadas de las técnicas de aprendizaje automático en la gestión de marketing son el comportamiento del consumidor, la optimización de la estructura del mercado de productos, la gestión de la mezcla de mercado, el marketing estratégico y el dominio de las finanzas, entre otros. Describen la sinergia entre el marketing y los sistemas inteligentes, especialmente las técnicas de aprendizaje automático.

En cuanto a la evolución del marketing digital, cuyo uso fue potenciado en el entorno de la crisis sanitaria covid-19, desde tiempos previos a la crisis los investigadores Bayoude et al. (2018) sostenían que el marketing digital está en constante evolución, se introducen regularmente nuevas herramientas con los nuevos hábitos de los consumidores y la multiplicación de datos, lo que a menudo obliga a los especialistas en marketing a profundizar en demasiados datos que pueden incluso no brindarles

la descripción general que necesitan para tomar decisiones comerciales. A partir de la revolución de la tecnología de aprendizaje automático en otras aplicaciones del mundo real, el aprendizaje automático ha ido cambiando el panorama del marketing digital. Las organizaciones de marketing implementaron o expandieron el uso del aprendizaje automático en el año 2018. Se vuelve más fácil predecir y analizar al comportamiento del consumidor con gran precisión. En su artículo, los investigadores propusieron un estado del arte de los principales y más utilizados modelos de aprendizaje automático en las diversas estrategias de marketing digital; además, mostraron cómo las herramientas de aprendizaje automático se pueden utilizar a gran escala con fines de marketing mediante el análisis de conjuntos de datos extremadamente grandes. La forma en que el ML se integra en las prácticas de marketing digital les ayuda a comprender mejor a los consumidores objetivo y optimizar sus interacciones con ellos.

La aplicación de la inteligencia artificial (IA) en marketing tiene como objetivo seguir y predecir continuamente las próximas decisiones de compra de los consumidores objetivo y mejorar su «recorrido» de consumidor. Al respecto, MR (2021) sostiene que el poder de la IA se refleja en sus elementos centrales: big data, aprendizaje automático y soluciones potentes. Los macrodatos permiten a los especialistas en marketing agregar y segmentar grandes cantidades de datos con un trabajo manual mínimo. Al usar estos datos, estarán seguros de que entregarán el mensaje correcto a los individuos adecuados en circunstancias adecuadas, a través del canal de su elección. El aprendizaje automático (aprendizaje profundo) permite a los especialistas en marketing comprender y obtener conclusiones lógicas a partir de grandes colecciones de datos. Pueden predecir tendencias de consumo, realizar seguimientos y análisis cuantitativo de las compras de los consumidores y predecir el próximo comportamiento del consumidor; es decir, vivimos en una era en la que las máquinas realmente entienden el mundo de la misma manera que los humanos. Las máquinas pueden identificar fácilmente conceptos y temas en una variedad de datos, interpretar emociones y comunicaciones humanas y generar respuestas adecuadas a los consumidores. Pueden predecir fácilmente el comportamiento y las decisiones de los compradores y utilizar esos datos para resolver problemas en el futuro. En los próximos años, los especialistas en marketing pueden esperar un mayor impacto de la IA, a través de más búsquedas inteligentes, anuncios

más inteligentes, entrega de contenido refinado, confianza en robots, aprendizaje continuo, prevención de fraudes y violaciones de datos, análisis de sentimientos, reconocimiento de imagen y voz, pronóstico de ventas, reconocimiento de idioma, servicio al cliente predictivo, segmentación de clientes, etc. (MR, 2021). Además, es posible que en el futuro descubran una relación más intensa entre los especialistas en marketing y las máquinas de inteligencia artificial.

Impulsado por las circunstancias de la crisis sanitaria covid-19, no se puede negar que el mundo está cambiando rápidamente a una era digital. Las personas consumen más contenido digital a diario al pasar más tiempo en línea. Las herramientas digitales que utilizan desempeñan un rol cada vez más significativo en sus vidas. Las empresas de marketing inteligentes reconocen esto e incorporan el concepto de inteligencia artificial a sus estrategias de marketing. Thiraviyam (2018) afirma que la inteligencia artificial está revolucionando el universo de actuación del marketing en la actualidad. Los macrodatos y el crecimiento exponencial de la potencia informática han abierto las puertas para el despegue de la inteligencia artificial. A medida que se sigue avanzando en el machine learning, las redes neuronales y la tecnología de aprendizaje profundo, más empresas recurren a la inteligencia artificial para hacer sus operaciones más inteligentes y eficientes. Cada vez más, el mundo del marketing está adoptando estos recursos para dar sentido a los datos, aprender más sobre los clientes y optimizar las operaciones. Este documento presenta una introducción sobre el marketing de inteligencia artificial y explora más a fondo cómo los especialistas en marketing están utilizando el poder de la inteligencia artificial, además, analiza la necesidad de utilizar esta estrategia para el marketing de productos y servicios.

El aprendizaje automático (ML) promete un gran valor para las aplicaciones relacionadas con el marketing. Sin embargo, la proliferación de tipos de datos, métodos, herramientas y lenguajes de programación dificulta la integración del conocimiento entre los equipos de análisis de marketing, lo que dificulta la colaboración. Al respecto, Villarroel y Silipo (2021) afirman que la programación visual puede facilitar la orquestación de proyectos de ML de una manera visual más intuitiva. En cuanto a la estrategia de marketing, Huang y Rust (2022) aseveran que la inteligencia artificial mecánica se puede utilizar para la segmentación (reconocimiento de segmentos), la inteligencia artificial pensante para la orientación (recomendación de segmento)

y sensación de inteligencia artificial para posicionamiento (resonancia de segmento).

De la revisión de la literatura expuesta líneas arriba se tiene que la segmentación de mercado y de clientes es una tarea fundamental para la gestión del marketing empresarial porque está relacionada con la identificación de grupos de consumidores que tiene un correlato con los productos elaborados por una empresa. Durante la investigación de mercado se recoge información sobre la percepción de los consumidores, utilizando en algunos casos recursos estadísticos que proporcionan pronósticos con un nivel de probabilidad de éxito y un grado de significancia estadística; sin embargo, los recursos de inteligencia artificial aportan al modelamiento estadístico el grado de certeza en los pronósticos.

El modelamiento de Machine Learning se realiza por medio de modelos matemáticos, econométricos y estadísticos. Para el caso en cuestión se empleó el modelo k-means, una herramienta cuantitativa utilizada para la conformación de conglomerados que fracciona el universo n de observaciones en k grupos diferentes guiados por criterios de afinidad interna y diferencia externa. En la Figura 1,

se ilustra el modelo k-means obtenido con el soporte de Jupyter Notebook Anaconda 3.0, que busca incluir en un grupo específico al vecino más próximo. Este algoritmo pertenece al grupo de algoritmos no supervisados. La formación de grupos de las n observaciones en los k grupos diferentes se efectúa minimizando la sumatoria de distancias en cada observación y el centroide del conglomerado.

De acuerdo con los objetivos del estudio y el análisis de los casos de identificación de grupos de conglomerados, los procesos de segmentación con el algoritmo k-means buscan que la varianza dentro de un clúster sea mínima, lo que se denomina búsqueda de homogeneidad interna. Por un lado, es fundamental encontrar grupos homogéneos que puedan satisfacerse con un determinado *marketing mix* y, por otro lado, se desea una amplia heterogeneidad externa entre los diferentes grupos de una población; por tanto, se premiará la selección de varianzas más altas entre grupos de conglomerados. En la Figura 1, se observa la presencia de 5 grupos y cada uno tiene componentes que conforman el conglomerado. El análisis de conglomerados es decisivo en el marketing porque está orientado a identificar la conducta del consumidor, hallar

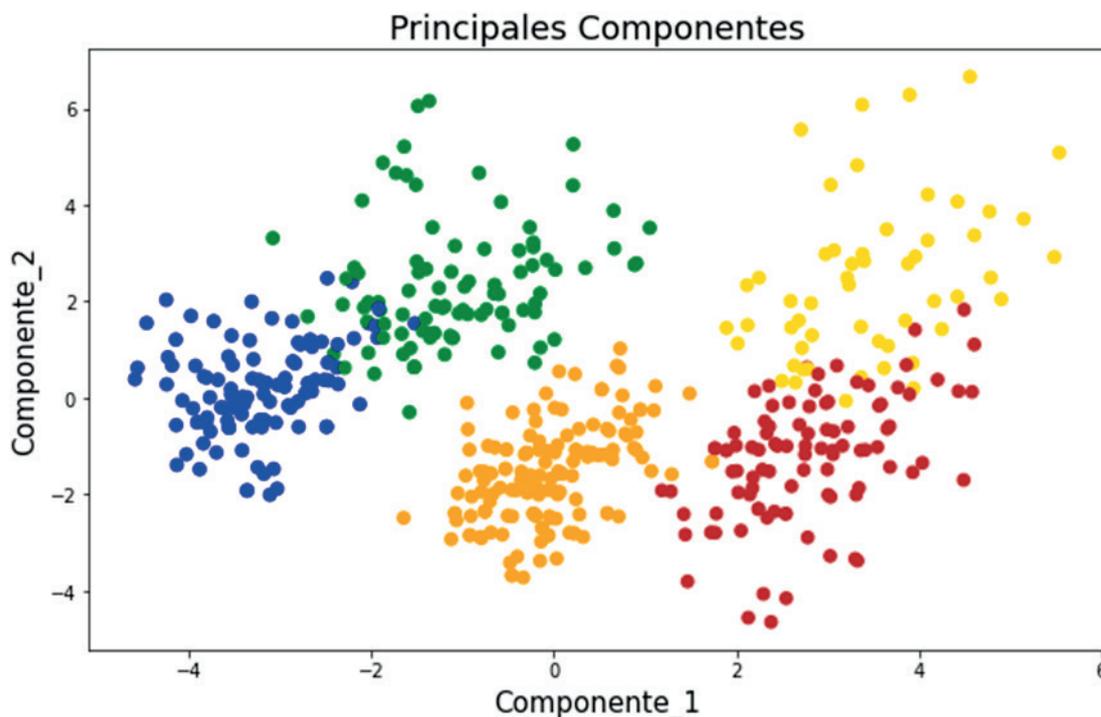


Figura 1. Clustering con modelo k-means.

Fuente: Elaboración propia.

oportunidades de productos nuevos, elegir mercados de prueba y decantar datos.

La Figura 2 muestra la ruta de formación de los conglomerados según la medida de distancia utilizada. El número de conglomerados depende de dónde se corte el dendrograma, por lo tanto, la decisión sobre el número óptimo de conglomerados es subjetiva. Es conveniente elegir un número de conglomerados que se sepa interpretar, que dependerá del número de conglomerados que se desean tener, por ejemplo, conglomerados de alto rendimiento, conglomerados de alta disponibilidad y conglomerados de alta eficiencia.

En estudio se analizan las implicancias de la adopción de tecnologías de inteligencia artificial para la identificación de mejores modelos de segmentación de mercado y clientes, temas de singular importancia para las empresas dedicadas a la gestión de mercados y de clientes. Los modelos cuantitativos que describen el comportamiento de las personas en el proceso de decisión de sus compras son del tipo multivariante, lo que quiere decir que en la decisión de compra de las personas intervienen

diversos factores cualitativos. Un factor importante en las decisiones de compra que esgrimen los individuos es el conjunto de manifestaciones de gustos, preferencias y estilos de vida de las personas. La granulometría de estas manifestaciones de gustos y preferencias y el avance de la tecnología han permitido a las personas disponer de información de manera casi instantánea, lo que ha provocado la existencia de pequeños segmentos de grupos de consumidores con determinados perfiles de consumo y cuyas exigencias antes de tomar la decisión de compra de un determinado producto se tienen en cuenta.

La presente investigación exhibe la importancia de la adopción de modelos de inteligencia artificial (machine learning) en la implementación de modelos de segmentación aplicados por las empresas para la gestión estratégica de sus clientes y mercados, con mayor razón en estos tiempos de covid-19 para evaluar permanentemente la dinámica de cambios en los gustos y preferencias de los consumidores, porque permite evaluar la eficiencia de la aplicación a través de un indicador de medición de performance, que en el caso específico

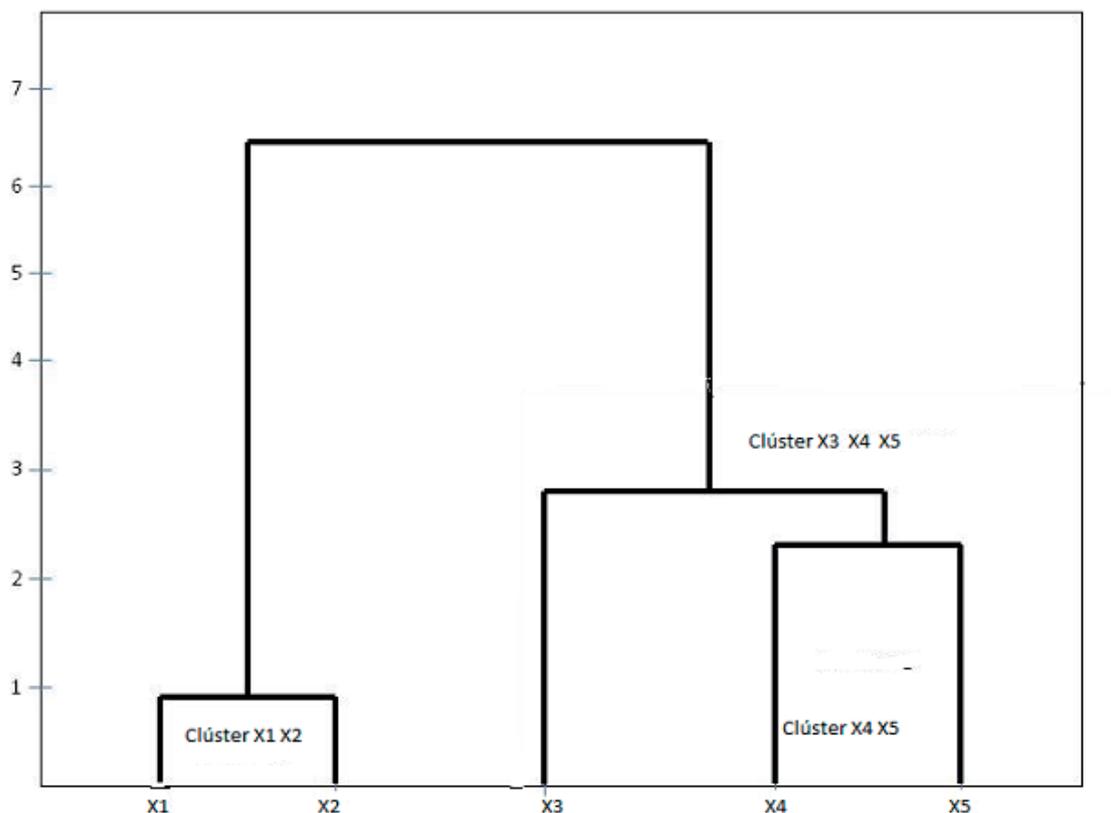


Figura 2. Representación gráfica de segmentación jerarquizada (dendrograma).

Fuente: Elaboración propia con el soporte de Anaconda 3.0.

del presente estudio fue el índice de Davies-Bouldin. Sin embargo, la implementación de modelos de clustering con herramientas de inteligencia artificial puede ser aplicada en todos los entornos empresariales y sociales en donde exista la necesidad de identificar grupos altamente diferenciados que muestren atributos significativos de cohesión interna, cuyos resultados sirvan para implementar políticas dinámicas, proactivas e innovadoras en la gestión organizacional.

Por todo lo anterior, la hipótesis de la investigación es la siguiente: La adopción de tecnologías de inteligencia artificial tiene una influencia positiva en las prácticas de segmentación de clientes en la investigación de mercados validado por el grado de certeza en los resultados de segmentación.

METODOLOGÍA

El estudio es del tipo analítico y sigue un diseño no experimental. Se elaboró un cuestionario de 6 preguntas con una escala de 7 niveles de respuestas con la opción de seleccionar una sola alternativa en cada pregunta del cuestionario. La encuesta relacionada con el análisis de conglomerados estuvo dirigida al público que asiste a los centros comerciales para realizar sus compras, con el propósito de obtener información sobre el sentimiento que les produce ir de compras, para cuyo efecto se identificaron 6 variables de actitud. Se solicitó a los consumidores que expresaran su grado de acuerdo con los enunciados que se describen a continuación:

V1: Realizar compras es muy divertido.

V2: Realizar compras no es bueno para mi presupuesto.

V3: Cuando voy de compras, aprovecho la oportunidad para comer fuera de casa.

V4: Cuando voy de compras, busco las mejores ofertas.

V5: No me motiva ir de compras.

V6: Puedo ahorrar mucho dinero cuando tengo la ocasión de comparar precios.

Cada una de las preguntas de la encuesta tuvo siete niveles de respuesta, en donde el nivel de mínima valoración fue 1 (totalmente en desacuerdo), y el nivel de máxima valoración fue 7 (totalmente de acuerdo).

Por conveniencia, la muestra estuvo compuesta por 506 personas mayores de 18 años, de ambos sexos. La muestra en mención no es del tipo aleatorio. La encuesta se suministró a través de Google Forms y estuvo disponible en la red durante un periodo de 15 días en el mes de marzo de 2022. Se seleccionó la prueba de Guttman para el análisis de fiabilidad con un índice de 0.659 y el índice de validez con prueba de esfericidad de Bartlett mediante el indicador de X^2 con p -valor < 0.05 a 95% de significancia estadística.

Una vez recogida la data mediante el formato de la encuesta, se preparó el dataset con los datos recogidos y se organizó en una hoja de Excel para luego exponer al modelo de clustering con RapidMiner Studio 9.10, siguiendo el proceso que se presenta en la Figura 3. Los ensayos de segmentación se ejecutaron según el procedimiento de segmentación aplicando modelo de aprendizaje no supervisado k-means. En primer lugar, se cargó la data de 506 registros, luego se sometió la data al proceso de normalización, para luego activar el clustering y, finalmente, someter los resultados a la evaluación del performance.

La Figura 3 describe el proceso secuencial de la aplicación de clustering. Primero, se importa la base de datos al ambiente de RapidMiner; se normaliza cada una de las variables porque el modelamiento para ML requiere de datos que tengan una distribución gaussiana; se aplica el modelo de clasificación de clúster con el que se obtienen los resultados para diferentes grupos en busca de la



Figura 3. Aplicación de modelamiento de Clustering con RapidMiner Studio 9.10

Fuente: Elaboración propia con RapidMiner Studio 9.10.

menor o la mayor distancia entre los grupos de personas; y, finalmente, los resultados se someten a la evaluación de performance para encontrar el número óptimo de grupos que tenga el mejor rendimiento.

El método utilizado para la ejecución de clústeres con RapidMiner 9.10, tal cual se presentó en la Figura 3, consiste en el análisis de distancia euclidiana al cuadrado entre dos individuos, que se define como la suma de cuadrados de las diferencias de todas las coordenadas de dos puntos, tal y como se muestra en la ecuación siguiente:

$$d(i, j) = \sqrt{\sum_k (X_{ik} - X_{jk})^2},$$

Donde, X_{ik} y X_{jk} son los valores individuales de cada uno de los puntos y el índice de Davies-Bouldin se utiliza para la evaluación del performance de los resultados. El modelo de Davies-Bouldin es un recurso utilizado para la validación interna de clustering, que se expresa de la siguiente manera:

$$DB = \left(\frac{1}{K}\right) \sum_{i=1}^k \max\left(\frac{\sigma_i + \sigma_j}{d(C_i, C_j)}\right)$$

Donde k es el número de clústeres, σ_i es la distancia promedio entre cada punto en el clúster i y el centroide del clúster, σ_j es la distancia promedio entre cada punto del clúster j y el centroide del clúster, y $d(C_i, C_j)$ es la distancia entre los centroides de los clústeres.

El algoritmo de clustering k-means es un algoritmo no supervisado que tiene muy buena escalabilidad para el procesamiento de datos. Es necesario que se especifique el número de grupos que se quiere formar para utilizar este algoritmo. Es posible encontrar el número óptimo de clústeres k , que es el principal objetivo de la investigación, vía ensayos.

RESULTADOS

El comportamiento descriptivo de cada una de las variables se exhibe en la Figura 4. Las respuestas

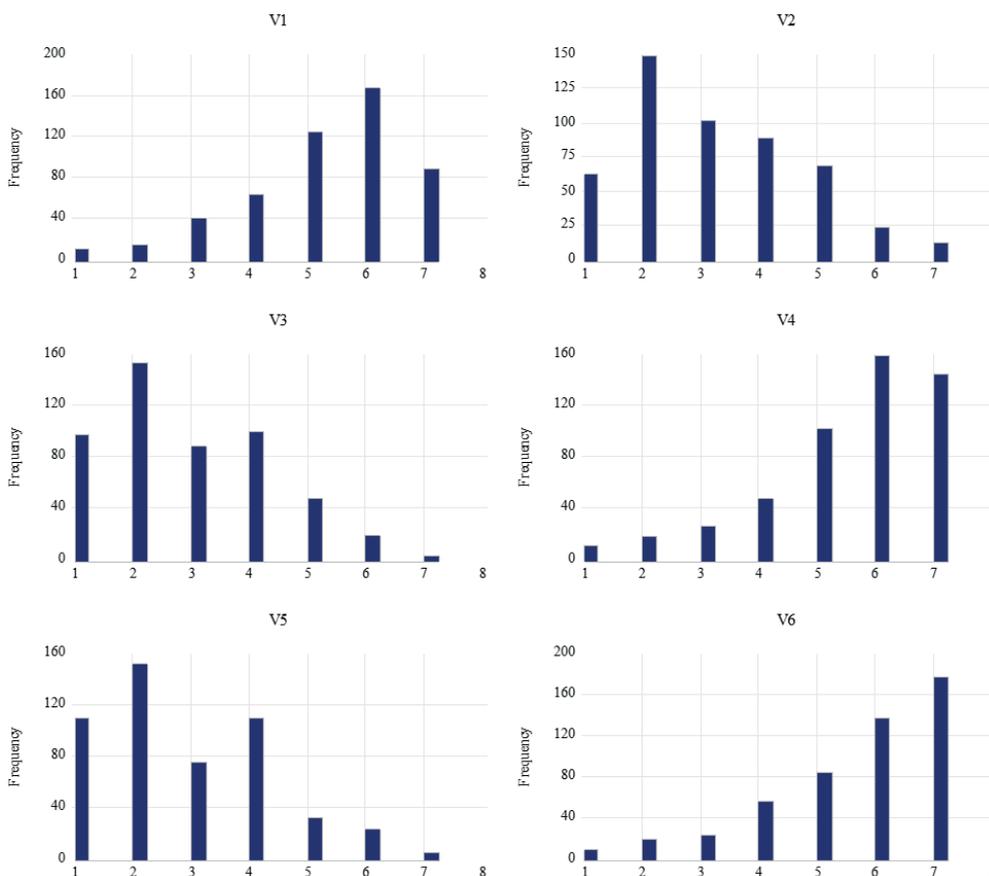


Figura 4. Distribución de las respuestas a los enunciados de segmentación.

Fuente: Elaboración propia con los datos de la encuesta.

de los encuestados fueron mayoritariamente positivas para todos los enunciados; se destaca que ir de compras les parece divertido y les provoca comparar precios, es una ocasión para aprovechar las ofertas y oportunidad para procurar ahorros en el presupuesto personal de los consumidores.

En la Tabla 1 se muestran los indicadores descriptivos obtenidos a partir de los resultados de la encuesta, que corresponden a la distribución de las respuestas brindadas por los encuestados a cada uno de los enunciados. La variable V5, «No me motiva el ir de compras», es la que tiene el valor promedio más bajo entre los valores de las 6 variables. Tal como se puede apreciar, se trata de un enunciado negativo al que los encuestados responden dando el valor mínimo; es decir, las respuestas brindadas por los encuestados resultan ser positivas. Por otro lado, una mayor valoración positiva corresponde al enunciado de la variable V6, «Puedo ahorrar mucho dinero cuando tengo la ocasión de comparar precios», pues es natural que los encuestados valoren disponer de la mayor cantidad de información posible para tomar la decisión de compra.

En la matriz de correlación de la Tabla 2, se observan relaciones directas entre las variables V1, V4 y V6, en concordancia con la ilustración descriptiva de las distribuciones mostradas en la Figura 4; a su vez, estas tres variables tienen relaciones indirectas con las variables V2, V3 y V5. En la matriz, se identifican las correlaciones más altas, concretamente, la relación que existe entre las variables V4 y V6 con $r=0.5721$ y las mínimas, como la que existe entre las variables V1 y V5 con $r=-0.3451$. Esta información es de gran importancia para la conformación de conglomerados en el proceso de construcción de grupos.

El indicador de centroide entre los conglomerados es la distancia entre los conglomerados. Los centroides mostrados en la Tabla 3 se obtuvieron con ensayos de segmentación con $k=1$, $k=2$, $k=3$, $k=4$ y $k=5$. Los centroides con valores altos marcan la heterogeneidad entre los grupos formados vía segmentación. Teniendo en cuenta el valor elevado de los centroides, el ensayo con $k=2$ es el que exhibe una pronunciada heterogeneidad respecto a otros grupos. En cuanto a las variables, V2, V3 y V5 son las que tienen centroides menores, y

Tabla 1. Indicadores descriptivos de la muestra.

Indicador	V1	V2	V3	V4	V5	V6
Media	5.2253	3.1482	2.8320	5.4921	2.7846	5.5850
Mediana	6.0000	3.0000	3.0000	6.0000	2.0000	6.0000
Máximo	7.0000	7.0000	7.0000	7.0000	7.0000	7.0000
Mínimo	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Desv. Estándar	1.4144	1.5142	1.4300	1.4762	1.4702	1.4971
Sesgo	-0.8732	0.5118	0.5245	-1.1214	0.6147	-1.0996
Curtosis	3.4038	2.5153	2.4708	3.8242	2.6413	3.6224
Jarque-Bera	67.7465	27.0447	29.1061	120.3733	34.5784	3.6224
Probabilidad	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Observaciones	506	506	506	506	506	506

Fuente: Elaboración propia con los datos de la encuesta.

Tabla 2. Matriz de correlación incondicional de la muestra.

Pregunta	V1	V2	V3	V4	V5	V6
V1	1.0000	-0.0711	0.0657	0.4257	-0.3451	0.3935
V2	-0.0711	1.0000	0.0408	0.0081	0.2883	0.0571
V3	0.0657	0.0408	1.0000	-0.0227	0.1297	0.0257
V4	0.4257	0.0081	-0.0227	1.0000	-0.1354	0.5721
V5	-0.3451	0.0081	0.1297	-0.1354	1.0000	-0.0496
V6	0.3935	0.2883	0.0256	0.5721	-0.0490	1.0000

Fuente: Elaboración propia con los datos de la encuesta.

en el otro extremo, V1, V4 y V6 exhiben centroides amplios entre los grupos formados.

La Tabla 4 muestra el resumen del análisis de varianza con el estadístico F univariante para cada una de las variables incluidas en el análisis. Se obtiene tomando los grupos definidos por los conglomerados como factor y cada una de las variables actuando en el análisis como variable dependiente. Se realiza el análisis de varianza tomando en cuenta como factor la clasificación de los grupos; en este caso, los 5 grupos pasan a ser los niveles de factor, lo que significa rechazar la igualdad de las medias de los grupos y aceptar la existencia de heterogeneidad entre grupos y homogeneidad dentro de los grupos. De este modo, se validan los cinco grupos organizados con el algoritmo k-means utilizando el RapidMiner 9.10 que confirma la consistencia estadística de la homogeneidad intragrupal y heterogeneidad intergrupala, dado el criterio p valor al 95% de confianza.

Los indicadores de performance para cuatro ensayos se presentan en la Tabla 5. El conglomerado que exhibe el menor valor DB en la medición de performance del modelo corresponde al ensayo que tiene $k=4$, con un porcentaje de performance de 24.30%, seguido de $k=5$ con 25.40% de performance. Por tanto, teniendo en cuenta el índice de Davies-Bouldin (DB), el número óptimo de

conglomerados es $k=4$, con 196 componentes en el grupo 1, 56 en el grupo 2, 133 en el grupo 3 y 121 en el grupo 4, que suman en total 506 encuestados, tal como se ilustra en la Figura 5.

Por otro lado, la prueba de Bartlett se ha utilizado como una herramienta auxiliar para el análisis de sedimentación que permite estimar la cantidad de grupos que pueden conformarse a partir de la data en estudio. Este método utiliza la distancia media de las observaciones a su centroide; se fija sustancialmente en la homogeneidad interna de un grupo. Cuanto mayor sea la cantidad de grupos (k), la varianza interna de un conglomerado tiende a disminuir, es decir, cuanto menor sea la distancia interna entre los elementos de un conglomerado tanto mejor, ya que los grupos son más homogéneos internamente.

La contrastación de la hipótesis de investigación se confirma con el indicador de Davies-Bouldin (Tabla 5), con el cual se realiza la validación interna de los grupos conformados. El mínimo índice DB obtenido de entre cuatro opciones de agrupación corresponde al segmento que tiene un valor de 24.30%, que es la opción óptima de segmentación, cuyo grado de certeza es proporcionado por el machine learning, con lo que queda demostrada la hipótesis de investigación.

Tabla 3. Indicadores de centroide de los conglomerados conformados con la muestra.

Variable	Clúster 1	Clúster 2	Clúster 3	Clúster 4	Clúster 5
V1	5.7566	3.9570	6.1053	3.4074	5.6022
V2	1.9934	4.2473	4.6579	2.7963	2.2903
V3	1.8421	3.0323	2.8947	2.3333	4.4624
V4	6.1184	5.2258	6.2456	2.7778	5.3871
V5	2.0461	4.7312	2.4035	2.3333	2.7742
V6	6.1579	5.5376	6.3158	2.9074	5.3548

Fuente: Elaboración propia con los datos de la encuesta.

Tabla 4. Análisis de varianza (ANOVA).

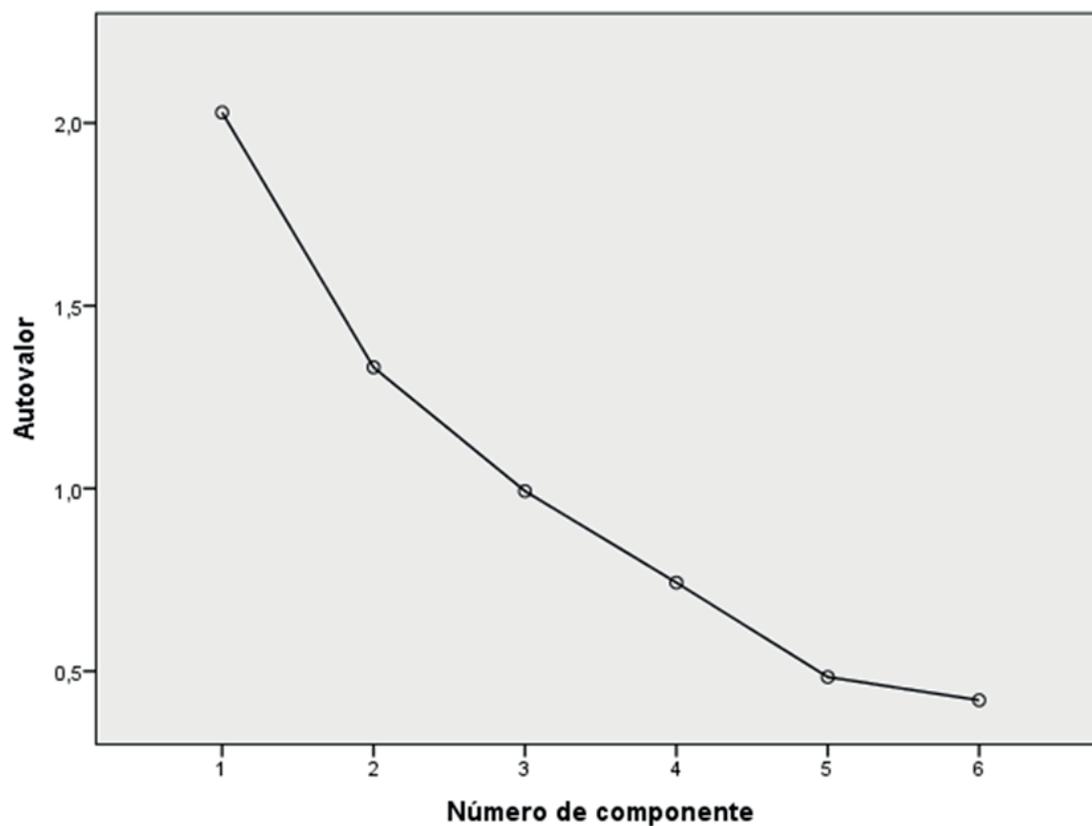
Variable	Clúster media cuadrática	gl	Error media cuadrática	gl	F	Sig.
V1	76 457.00	5	1.256	500	60.871	.000
V2	128 826.00	5	1.028	500	125.376	.000
V3	105 626.00	5	1.009	500	104.665	.000
V4	116 132.00	5	1.040	500	111.706	.000
V5	100 875.00	5	1.174	500	85.903	.000
V6	105 712.00	5	1.213	500	87.180	.000

Fuente: Elaboración propia con los datos de la encuesta.

Tabla 5. Medida de eficiencia de segmentación k-means.

N.º segmentos	Conglomerados	Componentes	Davies-Bouldin
2	Clúster 1	333	29.00%
	Clúster 2	173	
	Total:	506	
3	Clúster 1	212	27.90%
	Clúster 2	107	
	Clúster 3	187	
	Total:	506	
4	Clúster 1	196	24.30%
	Clúster 2	56	
	Clúster 3	133	
	Clúster 4	121	
	Total:	506	
5	Clúster 1	152	25.40%
	Clúster 2	93	
	Clúster 3	114	
	Clúster 4	54	
	Clúster 5	93	
	Total:	506	

Fuente: Elaboración propia con los datos de la encuesta.

**Figura 5.** Análisis de sedimentación para localizar la cantidad de grupos.

Fuente: Elaboración propia con los datos de la encuesta.

DISCUSIÓN

Para una empresa que actúa en diversos mercados ofertando sus productos, la actividad de estudios sobre el tema de grupos de consumidores es de mucha utilidad para la gestión estratégica de sus productos, así como para la gestión de mercados y clientes. Esta actividad, desde la perspectiva de investigación de mercados, se centra en recoger las percepciones de gustos y preferencias de los consumidores en relación con determinados productos que circulan en el mercado. Al disponer de la información oportuna sobre los gustos y preferencias de los consumidores, las empresas pueden alinear sus ofertas rápidamente a través de sus departamentos de innovación y así orientar determinados productos a grupos específicos de consumidores.

La gestión de relaciones con los consumidores tiene como objetivo construir relaciones mediante la segmentación de clientes y el diseño de herramientas de marketing adecuadas. Las empresas de marketing inteligentes reconocen esto e incorporan el concepto de inteligencia artificial a sus estrategias de marketing. En la actualidad, la inteligencia artificial está revolucionando el mundo del marketing (Thiraviyam, 2018).

La ciencia del marketing tiene una larga tradición de aceptación de nuevos desafíos, nuevos métodos y nuevas disciplinas. Por ejemplo, la adopción de tecnologías computacionales para las actividades de investigación de mercados, así como el uso del algoritmo de clustering como otra alternativa de aplicación de inteligencia artificial (IA) para la identificación de grupos de consumidores. En la actualidad, la ciencia del marketing se basa en los diversos esfuerzos de los investigadores que, durante casi 50 años, han sintetizado soluciones de una variedad de disciplinas para proporcionar una nueva perspectiva para abordar los problemas de marketing. La mayoría de las veces, el crisol de la ciencia del marketing ha proporcionado a otras disciplinas modelos y métodos que son mejores y más sólidos.

El progreso reciente en el aprendizaje automático, conocido con el nombre de machine learning, ha sido impactado positivamente tanto por el acelerado desarrollo de nuevos algoritmos y teorías de aprendizaje, como por la explosión dinámica en la accesibilidad a datos en línea y la disponibilidad de recursos informáticos de bajo costo. La adopción de métodos de aprendizaje automático acompañado por el uso intensivo de datos puede encontrarse en todas las expresiones de la ciencia, la tecnología y el comercio, lo que en el campo del marketing empresarial conduce a los responsables de la toma

de decisiones a basarse más en la evidencia que en la intuición (Jordan y Mitchell, 2015).

El marketing aprovecha los beneficios de la IA y los modelos de aprendizaje automático para automatizar, optimizar y aumentar el accionar de transformación de los datos en acciones e interacciones con el alcance de predecir comportamientos, anticipar necesidades e hiperpersonalizar mensajes (Kaličanin et al., 2019). El aprendizaje automático ofrece un gran valor potencial para las aplicaciones relacionadas con el marketing (Villarroel y Silipo, 2021).

La aplicación de la segmentación por conglomerados con la adopción de tecnologías de inteligencia artificial en la gestión de marketing digital, tal cual afirma Brei (2020), facilita la toma de decisiones relacionados con la gestión de clientes y productos. Por su parte, Ullal et al. (2021) afirman que la adopción de herramientas de IA tiene como objetivo brindar a las empresas información en tiempo real. Asimismo, Jarek y Mazurek (2019) hacen notar que el machine learning combina la ciencia de datos, los métodos cuantitativos y la programación computacional.

La IA contribuye al desarrollo personalizado de estrategias de marketing, creando estrategias de comunicación adaptadas a los clientes según sus gustos y preferencias, sus datos de geolocalización y toda la información que se requiera conocer de ellos. Además, soporta la segmentación de datos de clientes, clasificando grupos de consumidores según sus características. Esto permite a las empresas evaluar de manera continua, con perspectiva de corto plazo, la evolución de los clientes en el tiempo y gestionar las estrategias de manera dinámica en el camino. Por otro lado, también permite a las empresas cazar oportunidades en ventas y, a partir de la información proporcionada por la segmentación de mercados, observar en detalle los cambios en las preferencias de los consumidores. También se pueden identificar las oportunidades de negocio teniendo en cuenta las visitas a un sitio de productos o a partir de las compras que los consumidores han realizado en el transcurso del tiempo a través de medios digitales, lo que hoy se conoce como analítica web.

CONCLUSIONES

La segmentación es muy útil en el marketing empresarial porque permite tomar decisiones en tiempo real, evitar pérdidas de clientes y realizar seguimientos a tendencias de gustos y preferencias de los consumidores.

Asimismo, gracias a los datos, se puede observar cómo las personas desarrollan sentimientos hacia las marcas de la empresa, ya sean manifestaciones positivas o negativas. La crisis sanitaria covid-19 ha provocado cambios sustanciales en el comportamiento de consumidores, los que han impactado sobre su forma de comprar y tomar decisiones de compra.

En el paradigma de clustering que permite encontrar la distancia euclídea entre los grupos de consumidores en el ámbito del marketing empresarial, las distancias mínimas indican que no existen marcadas divergencias en los gustos y preferencias de los consumidores, mientras que las distancias grandes denotan que existen amplias diferencias.

A partir de esa información, se identificaron 4 grupos de consumidores con percepciones marcadamente diferentes respecto a realizar compras: un primer grupo que expresa ahorrar en periodos de ofertas, un segundo grupo que prefiere disponer de una amplia gama de ofertas y precios para comprar un producto de su preferencia, un tercer grupo que manifiesta que cuando sale de compras aprovecha comer fuera de casa con la familia y un cuarto grupo minoritario que manifiesta su no preferencia de realizar compras porque no es bueno para su presupuesto y que no le motiva ir de compras.

En conclusión, se puede afirmar que en este estudio se demostró la utilidad de inteligencia artificial para la estimación del número óptimo de segmentos, sustentado por el indicador de Davies-Bouldin basado en la performance y la similitud de los conglomerados, con lo que quedó demostrada la hipótesis de investigación.

AGRADECIMIENTO

A mi Universidad Nacional Jorge Basadre y a su Escuela de Ingeniería Comercial.

REFERENCIAS

- [1] Abakouy, R., En-Naimi, E. M., El Haddadi, A., y Lotfi, E. (2019). Data-driven marketing: How machine learning will improve decision-making for marketers. *SCA '19: Proceedings of the 4th International Conference on Smart City Applications*. <https://doi.org/10.1145/3368756.3369024>
- [2] Baltes, P. L. (2015). Content marketing - the fundamental tool of digital marketing. *Bulletin of the Transilvania University of Brasov*. 8(2), 111-118.
- [3] Bayoude, K., Ouassit, Y., Ardchir, S., y Azouazi, M. (2018). How machine learning potentials are transforming the practice of digital marketing: State of the art. *Periodicals of Engineering and Natural Sciences*, 6(2), 373-379. <https://doi.org/10.21533/pen.v6i2.526>
- [4] Brei, V. A. (2020). Machine learning in marketing: Overview, Learning Strategies, Applications, and Future Developments. *Foundations and Trends in Marketing*, 14(3), 173-236. <https://doi.org/10.1561/17000000065>
- [5] Casco, A. R. (2020). Efectos de la pandemia de COVID-19 en el comportamiento del consumidor. *Innovare: Revista de Ciencia y Tecnología*, 9(2), 98-105. <https://doi.org/10.5377/innovare.v9i2.10208>
- [6] Dzyabura, D., y Hauser, J. R. (2011). Active machine learning for consideration heuristics. *Marketing Science*, 30(5), 801-819. <https://doi.org/10.1287/mksc.1110.0660>
- [7] Eger, L., Komárková, L., Egerová, D., y Mičík, M. (2021). The effect of COVID-19 on consumer shopping behaviour: Generational cohort perspective. *Journal of Retailing and Consumer Services*, 61. <https://doi.org/10.1016/j.jretconser.2021.102542>
- [8] Ellickson, P. B., Kar, W., y Reeder, J. C. (2022). Estimating Marketing Component Effects: Double Machine Learning from Targeted Email Promotions. *Marketing Science*. <https://doi.org/10.1287/mksc.2022.1401>
- [9] Huang, M. H., y Rust, R. T. (2022). A Framework for Collaborative Artificial Intelligence in Marketing. *Journal of Retailing*, 98(2), 209-223. <https://doi.org/10.1016/j.jretai.2021.03.001>
- [10] Jain, S., y Kain, M. (2018). Prediction for Stock Marketing Using Machine Learning. *International Journal on Recent and Innovation Trends in Computing and Communication*, 6(4), 131-135.
- [11] Jarek, K., y Mazurek, G. (2019). Marketing and Artificial Intelligence. *Central European Business Review*, 8(2), 46-55. <https://doi.org/10.18267/j.cebr.213>
- [12] Jordan, M. I., y Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245), 255-260. <https://doi.org/10.1126/science.aaa8415>
- [13] Kabir, M. Y., y Madria, S. (2021). EMOCOV: Machine learning for emotion detection,

- analysis and visualization using COVID-19 tweets. *Online Social Networks and Media*, 23. <https://doi.org/10.1016/j.osnem.2021.100135>
- [14] Kaličanin, K., Čolović, M., Njeguš, A., y Mitić, V. (2019). Benefits of Artificial Intelligence and Machine Learning in Marketing, *Sinteza 2019 - International Scientific Conference on Information Technology and Data Related Research*, 472-477. <https://doi.org/10.15308/sinteza-2019-472-477>
- [15] Kumar, A., Gawande, A., y Brar, V. (2020). Marketing Tactics in Times of Covid-19. *Vidyabharati International Interdisciplinary Research Journal*, 11(2), 263-266.
- [16] Mahajan, K. S., Jamsandekar, S. S., y Gurav, A. M. (2017). Machine Learning Approach for Marketing Intelligence: Managerial Application. *International Journal Of Engineering And Computer Science*, 6(7), 21929-21936.
- [17] Mahendiran, A., Saravanan, N., Subramanian, N., y Sairam, N. (2012). Implementation of K-means clustering in cloud computing environment. *Research Journal of Applied Sciences, Engineering and Technology*, 4(10), 1391-1394.
- [18] Mehta, S., Saxena, T., y Purohit, N. (2020). The New Consumer Behaviour Paradigm amid COVID-19: Permanent or Transient? *Journal of Health Management*, 22(2), 291-301. <https://doi.org/10.1177/0972063420940834>
- [19] Miklosik, A., Kuchta, M., Evans, N., y Zak, S. (2019). Towards the Adoption of Machine Learning-Based Analytical Tools in Digital Marketing. *IEEE Access*, 7, 85705-85718. <https://doi.org/10.1109/ACCESS.2019.2924425>
- [20] MR, A. (2021). Artificial Intelligence and Marketing. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, 12(4), 1247-1256. <https://doi.org/10.17762/turcomat.v12i4.1184>
- [21] Rodríguez-Rodríguez, I., Rodríguez, J. V., Shirvanizadeh, N., Ortiz, A., & Pardo-Quiles, D. J. (2021). Applications of Artificial Intelligence, Machine Learning, Big Data and the Internet of Things to the COVID-19 Pandemic: A Scientometric Review Using Text Mining. *International Journal of Environmental Research and Public Health*, 18(16). <https://doi.org/10.3390/ijerph18168578>
- [22] Thiraviyam, T. (2018). Artificial Intelligence Marketing. *International Journal of Recent Research Aspects*, 19(4), 449-452.
- [23] Ullal, M. S., Hawaldar, I. T., Soni, R., y Nadeem, M. (2021). The Role of Machine Learning in Digital Marketing. *SAGE Open*, 11(4). <https://doi.org/10.1177/21582440211050394>
- [24] Villarroel Ordenes, F., y Silipo, R. (2021). Machine learning for marketing on the KNIME Hub: The development of a live repository for marketing applications. *Journal of Business Research*, 137, 393-410. <https://doi.org/10.1016/j.jbusres.2021.08.036>
- [25] Yıldırım, M., Geçer, E., y Akgül, Ö. (2021). The impacts of vulnerability, perceived risk, and fear on preventive behaviours against COVID-19. *Psychology, Health and Medicine*, 26(1), 35-43. <https://doi.org/10.1080/13548506.2020.1776891>