



Artículo original / Original Article

Modelo basado en técnicas de minería de datos para la segmentación de clientes en empresas distribuidoras

Model based on data mining techniques for customer segmentation in distribution companies

Jaime Chacaliaza-Almeyda ¹; Richard Enrique Injante-Oré ¹

¹ Universidad Nacional de San Martín, Tarapoto, Perú

Recibido: 23/10/2024

Aceptado: 15/11/2024

Publicado: 30/01/2025

*Autor de correspondencia: richard@unsm.edu.pe

Resumen: Las empresas distribuidoras enfrentan desafíos críticos para optimizar estrategias comerciales en mercados competitivos debido a la gestión de múltiples categorías de productos y bases diversas de clientes. Esta investigación tuvo como objetivo diseñar e implementar un modelo integral basado en técnicas de minería de datos para la segmentación efectiva de clientes en una empresa distribuidora. Se empleó la metodología KDD integrando el análisis RFM con el algoritmo K-means y un enfoque de cuartiles. Se analizaron 44,000 registros transaccionales del período 10 meses de cinco sucursales, aplicando un modelo estructurado en siete pasos sistemáticos. Los resultados identificaron cinco segmentos distintos de clientes con características diferenciadas y patrones específicos en 25 categorías de productos. La validación mediante el método de silueta confirmó K-means con cinco clústeres como configuración óptima (coeficiente=0.4). El modelo desarrollado permitió identificar patrones complejos de compra y facilitó la implementación de estrategias comerciales personalizadas, mejorando el proceso de toma de decisiones.

Palabras clave: análisis RFM; algoritmos clustering; empresas distribuidoras; inteligencia artificial; segmentación clientes

Abstract: Distribution companies face critical challenges in optimizing business strategies in competitive markets due to the management of multiple product categories and diverse customer bases. This study aimed to design and implement a comprehensive model based on data mining techniques for effective customer segmentation in a distribution company. The KDD methodology was employed, integrating RFM analysis with the K-means algorithm and a quartile-based approach. A dataset of 44,000 transactional records covering a 10-month period from five branches was analyzed using a structured seven-step model. The results identified five distinct customer segments with differentiated characteristics and specific patterns across 25 product categories. Validation using the silhouette method confirmed K-means with five clusters as the optimal configuration (coefficient = 0.4). The developed model enabled the identification of complex purchasing patterns and facilitated the implementation of personalized business strategies, thereby improving the decision-making process.

Keywords: artificial intelligence; clustering algorithms; customer segmentation; distribution companies; RFM análisis

1. Introducción

La transformación digital ha revolucionado la manera en que las organizaciones interactúan con sus clientes y gestionan sus estrategias comerciales. En este contexto, la segmentación de clientes se configura como una herramienta estratégica esencial que permite comprender mejor la base de consumidores, optimizar recursos y maximizar el retorno de inversión en marketing (Zhang et al., 2023). Esta relevancia se acentúa particularmente en el sector de distribución, caracterizado por catálogos extensos de productos, múltiples puntos de venta y una base diversa de clientes con necesidades heterogéneas (Kumar & Shah, 2024).

El panorama competitivo actual presenta desafíos sin precedentes para las empresas distribuidoras. La masificación de datos transaccionales, junto con una creciente competencia y expectativas cada vez más sofisticadas de los clientes, ha configurado un escenario en el cual la diferenciación basada únicamente en precio o producto resulta insuficiente (Wang et al., 2023). Las organizaciones que logran ventajas sostenibles son aquellas capaces de interpretar con profundidad los patrones de comportamiento de sus clientes y transformar ese conocimiento en estrategias personalizadas y efectivas (Chen & Liu, 2024).

No obstante, en mercados emergentes muchas distribuidoras aún recurren a métodos empíricos y tradicionales para categorizar a sus clientes. Estos enfoques, sustentados frecuentemente en criterios unidimensionales como el volumen de compra o la localización geográfica, resultan limitados frente a la complejidad multifacética del consumidor contemporáneo (Martinez & Silva, 2024). Como consecuencia, se restringe la posibilidad de implementar estrategias diferenciadas, lo que deriva en pérdidas de oportunidades comerciales y en un uso ineficiente de los recursos de marketing y ventas.

Ante este escenario, la literatura científica ha documentado ampliamente el potencial de la minería de datos para robustecer los procesos de segmentación (Castillo-Rojas et al., 2017). El análisis RFM (Recencia, Frecuencia, Monto), introducido por Hughes (1994) y perfeccionado en décadas posteriores, se consolidó como un método eficaz para valorar clientes en función de su historial transaccional (Liu et al., 2024). Al combinarse con algoritmos modernos de clustering, esta técnica permite identificar grupos de clientes internamente homogéneos y heterogéneos entre sí, posibilitando la formulación de estrategias diferenciadas (Park & Kim, 2023).

Investigaciones recientes han validado la efectividad de estas aproximaciones en diversos sectores. Dursun & Caber (2016), por ejemplo, demostraron la utilidad del análisis RFM junto con mapas autoorganizados (SOM) y K-means en el sector hotelero, identificando ocho tipologías de huéspedes con patrones de consumo distintos. Sus resultados evidenciaron que un 15% de los clientes generaba el 45% de los ingresos y el 70% de las utilidades, lo que resalta la necesidad de identificar y gestionar segmentos de alto valor. De forma similar, Aryuni et al. (2018) compararon K-means y K-medoides en un contexto bancario, concluyendo que K-means ofrecía mejores resultados de acuerdo con los índices AWC (Average Within Cluster) y DBI (Davies-Bouldin Index).

A pesar de estos avances, la literatura aún presenta vacíos significativos en torno a modelos integrales aplicados al sector de distribución automotriz en mercados emergentes. Gran parte de las investigaciones se han centrado en banca, turismo o retail, sin contemplar la complejidad de empresas que gestionan amplias categorías técnicas de productos, operan en múltiples ubicaciones y atienden tanto a consumidores individuales como a clientes corporativos (Rodriguez et al., 2023). Esta brecha resulta especialmente notoria en América Latina, donde la informalidad del mercado, la volatilidad de la demanda y las limitaciones tecnológicas constituyen factores críticos (García & López, 2024).

Otro aspecto poco explorado es la integración de la dimensión de productos en el análisis de segmentación. En el sector de distribución automotriz, esta limitación es particularmente relevante, ya que la relación entre segmentos de clientes y categorías de productos puede revelar oportunidades estratégicas en ventas cruzadas, expansión de mercado y optimización de

inventarios (Thompson & Anderson, 2023). En este sentido, la capacidad de identificar no solo qué clientes aportan mayor valor, sino también qué productos específicos consumen, constituye una ventaja competitiva clave en mercados con márgenes ajustados.

El caso de las empresas distribuidoras del oriente peruano refleja claramente esta problemática. Estas organizaciones enfrentan una dispersión geográfica significativa, una amplia heterogeneidad en la sofisticación técnica de sus clientes y una base que incluye desde pequeños talleres mecánicos hasta grandes flotas corporativas (Wilson et al., 2023). Esta diversidad exige modelos de segmentación multidimensionales capaces de capturar y procesar simultáneamente diversas aristas del comportamiento del consumidor.

La empresa Suministros (nombre reservado) de este estudio constituye un ejemplo representativo. Con más de veinte años de operación y presencia en cinco ciudades estratégicas de la región, ha acumulado una vasta base de datos transaccionales que permanece subutilizada. A pesar de disponer de un sistema ERP que registra alrededor de 4,000 transacciones mensuales, la compañía sigue dependiendo de reportes básicos y segmentaciones empíricas que no reflejan la complejidad de su cartera de clientes. Esta situación restringe la efectividad de sus estrategias y genera una pérdida considerable de oportunidades competitivas basadas en el uso inteligente de datos.

El presente estudio busca atender estas brechas mediante el diseño de un modelo integral que combina la metodología KDD (Knowledge Discovery in Databases) con el análisis RFM y algoritmos de clustering avanzados. La principal innovación radica en la incorporación de un enfoque de cuartiles que no solo segmenta a los clientes, sino que también establece vínculos significativos entre segmentos y categorías de productos. Este enfoque bidimensional ofrece una visión más completa y operativa del mercado, lo cual facilita el diseño de estrategias comerciales diferenciadas y fundamentadas en evidencia. En consecuencia, el objetivo central de esta investigación fue diseñar, implementar y validar un modelo de minería de datos que optimice el proceso de segmentación de clientes en la empresa distribuidora.

2. Materiales y métodos

2.1. Área de estudio y contexto organizacional

La investigación se desarrolló en una empresa suministros, una organización líder en la distribución de productos automotrices con más de 20 años de trayectoria en el mercado peruano. La empresa opera una red estratégica de cinco sucursales ubicadas en ciudades clave del oriente del país: Tarapoto (sede principal), Pucallpa, Iquitos, Jaén y Tocache. Esta distribución geográfica le permite atender un mercado diverso que incluye talleres mecánicos independientes, concesionarios, flotas corporativas y consumidores finales.

El estudio se centró en el análisis exhaustivo de los datos transaccionales almacenados en el sistema ERP empresarial, que registra un promedio de 4,000 transacciones mensuales, totalizando aproximadamente 44,000 registros durante el período de análisis. La empresa maneja un catálogo de más de 5,000 SKUs agrupados en 25 categorías principales de productos automotrices, desde lubricantes y filtros hasta componentes especializados de motor y transmisión.

2.2. Tipo y diseño de investigación

Se realizó una investigación aplicada de nivel descriptivo-correlacional, empleando un diseño no experimental de corte transversal. Este enfoque metodológico permitió analizar los datos históricos de clientes sin manipular las variables de estudio, describir las características y patrones de comportamiento de los segmentos identificados, establecer relaciones entre segmentos de clientes y categorías de productos, y generar conocimiento aplicable directamente a la realidad empresarial. El período de análisis comprendió transacciones durante 9 meses,

seleccionado por representar un ciclo comercial completo sin disrupciones extraordinarias del mercado.

2.3. Metodología KDD y desarrollo del modelo

El desarrollo del modelo siguió rigurosamente la metodología KDD (Knowledge Discovery in Databases) propuesta por Fayyad et al. (1996), adaptada específicamente para el contexto de segmentación de clientes en empresas distribuidoras. El modelo propuesto integra siete pasos secuenciales y sistemáticos:

Paso 1: Determinación del objetivo Se estableció el objetivo estratégico de la segmentación y se definió la ponderación de las variables RFM. Para el caso de estudio, se seleccionó como objetivo principal "segmentar a los clientes por frecuencia de compras para las categorías de productos", asignando ponderaciones de Recencia (R): peso = 1, Frecuencia (F): peso = 2 (variable predominante), y Monto (M): peso = 1. Esta ponderación refleja la importancia estratégica de mantener una relación comercial activa y frecuente con los clientes, considerada fundamental en el sector de distribución automotriz.

Paso 2: Selección de datos Se identificaron y extrajeron los campos relevantes del sistema ERP empresarial. La selección se basó en la necesidad de capturar cuatro dimensiones fundamentales: temporal (fechas de transacciones), transaccional (identificadores únicos de documentos), económica (montos de venta), y de producto (categorías). Los campos seleccionados fueron identificador único del cliente, fecha de cada venta, categoría del producto vendido, monto de la transacción, y número y tipo de documento.

Paso 3: Limpieza y pre-procesado Se implementó un protocolo exhaustivo de limpieza de datos que incluyó eliminación de registros problemáticos (categorías de productos vacías, documentos no correspondientes, clientes del grupo comercial interno, transacciones promocionales con precio cero), estandarización de formatos (conversión de fechas al formato ISO, normalización de nombres de categorías, unificación de códigos de clientes duplicados), y validación de integridad (verificación de montos positivos, validación de fechas, confirmación de existencia de campos requeridos).

Paso 4: Transformación La transformación se desarrolló en dos procesos paralelos. El Proceso 1 construyó el perfil RFM de cada cliente calculando Recencia como días transcurridos desde la última compra hasta el final del período, Frecuencia como número total de transacciones realizadas, y Monto como suma del valor total de todas las compras. Para normalizar estas variables con escalas diferentes, se aplicó el enfoque de cuartiles, dividiendo a los clientes en cuatro grupos iguales para cada variable. El Proceso 2 construyó una matriz binaria cliente-producto que captura qué productos compra cada cliente.

Paso 5: Segmentación mediante clustering Se aplicó el algoritmo K-means sobre los perfiles RFM normalizados. La determinación del número óptimo de segmentos empleó tres métodos complementarios: método del codo (evaluando configuraciones de 2 a 10 segmentos), método de la silueta (midiendo separación entre grupos), y estadística de la brecha (comparando dispersión observada con la esperada por azar). Los tres métodos convergieron en $k=5$ como configuración óptima. Se aplicó ponderación estratégica con la fórmula $P = (R \times 1 + F \times 2 + M \times 1) / 4$ para priorizar la frecuencia de compra.

Paso 6: Construcción de la matriz comparativa Se integró la segmentación de clientes con el análisis de categorías de productos, creando una herramienta bidimensional única. Se contabilizó cuántos clientes de cada segmento compraron cada categoría y se aplicó nuevamente el enfoque de cuartiles para normalizar las diferencias entre segmentos de diferentes tamaños.

Paso 7: Interpretación estratégica Se realizó el análisis detallado de los resultados para generar recomendaciones estratégicas accionables, considerando características distintivas de cada segmento, patrones de consumo por categoría de producto, oportunidades de venta cruzada y desarrollo de mercado, y estrategias diferenciadas por segmento.

2.4. Técnicas e instrumentos de análisis

El procesamiento y análisis de datos se realizó utilizando R versión 4.0.3 con RStudio 1.4.1106. Las librerías principales incluyeron tidyverse para manipulación de datos, cluster para algoritmos de clustering, factoextra para visualización y validación, ggplot2 para gráficos avanzados, y corrplot para matrices de correlación.

2.5. Validación y comparación de métodos

Se implementó un protocolo exhaustivo de validación que incluyó validación del número de clústeres mediante WSS (Within Sum of Squares), coeficiente de silueta promedio, y estadística de la brecha con $B=100$ bootstrap samples. Se realizó comparación de algoritmos K-means vs CLARA, evaluación mediante índices internos (Silueta, Dunn, Davies-Bouldin), y análisis de estabilidad con diferentes semillas aleatorias.

2.6. Consideraciones éticas

La investigación se condujo siguiendo estrictos protocolos éticos incluyendo anonimización completa de datos de clientes, firma de acuerdo de confidencialidad con la empresa, uso exclusivo de datos para fines de investigación, y no divulgación de información comercial sensible.

3. Resultados y discusión

3.1. Construcción progresiva del modelo de segmentación

El desarrollo del modelo siguió una lógica constructiva donde cada paso agregó una capa de comprensión sobre los clientes. La transformación exitosa de 44,000 registros transaccionales en 2,700 perfiles únicos de clientes caracterizados por tres dimensiones RFM constituyó el primer resultado significativo. Los clientes mostraron una distribución de recencia bimodal con concentraciones en compras muy recientes (0-30 días) y un segundo grupo con compras antiguas (>180 días), sugiriendo dos poblaciones distintas: activos e inactivos. La distribución de frecuencia siguió una distribución de ley de potencia típica en retail, donde pocos clientes (15%) generaban la mayoría de transacciones (65%), validando la aplicabilidad del principio de Pareto. La distribución de monto presentó alta variabilidad con coeficiente de variación de 2.3, indicando la presencia de clientes con patrones de gasto muy diversos.

3.2. Descubrimiento de segmentos mediante clustering

La aplicación del algoritmo K-means sobre los perfiles RFM normalizados fue un proceso iterativo de descubrimiento. La búsqueda del número óptimo de segmentos reveló que con 5 segmentos se lograba el balance óptimo entre separación y manejabilidad (Christy et al., 2021). El coeficiente de silueta de 0.40 para 5 segmentos indicó una estructura de agrupamiento "justa" según los estándares de Kaufman & Rousseeuw (1990). Los cinco clústeres emergentes mostraron perfiles claramente diferenciados que, tras la aplicación de la ponderación estratégica, se transformaron en segmentos de negocio ordenados por valor .

Tabla 1. Transformación de clústeres a segmentos estratégicos

Clúster Original	R	F	M	Ponderado	Segmento	Interpretación de Negocio
Clúster 5	3.72	3.76	3.57	3.70	A	Los mejores clientes: compran frecuentemente, recientemente y por montos altos
Clúster 4	2.15	2.79	3.45	2.80	B	Clientes valiosos: compras de alto ticket pero menos frecuentes
Clúster 1	2.81	2.56	2.08	2.50	C	La base estable: desempeño consistente pero moderado
Clúster 2	3.51	1.26	1.56	1.90	D	Clientes nuevos o evaluando: compraron recientemente pero poco
Clúster 3	1.33	1.10	1.46	1.25	E	Clientes en riesgo: baja actividad en todas las dimensiones

3.3. Validación del principio de Pareto

Un hallazgo fundamental fue la confirmación del principio 80/20 en los datos reales, donde el Segmento A superó las expectativas: 20% de clientes generando 45% de ingresos (ratio 2.25x), mientras que el Segmento E confirmó la existencia de una "larga cola" de clientes de bajo valor.

Tabla 2. Distribución de valor por segmento

Segmento	Nº Clientes	% de la Base	% de Ingresos	Ratio Valor/Tamaño
A	542	20.1%	45.2%	2.25x
B	673	24.9%	28.3%	1.14x
C	486	18.0%	15.7%	0.87x
D	298	11.0%	7.2%	0.65x
E	701	26.0%	3.6%	0.14x
Segmento	Nº de clientes	% de la base	% de ingresos	Ratio Valor/Tamaño
Total	2,700	100%	100%	-

3.4. Análisis de patrones de consumo por categorías

Paralelamente a la segmentación RFM, se mapearon las preferencias de producto de cada cliente, revelando 25 categorías de productos con patrones de demanda muy diversos. La alta concentración en las primeras cuatro categorías (motores 4T, motores a gasolina, calcio, motores a diesel) confirmó el core business de la empresa.

3.5. Creación de la matriz comparativa

La innovación clave fue aplicar el enfoque de cuartiles por segunda vez, transformando números absolutos en posiciones relativas dentro de cada segmento. Este proceso se repitió para cada segmento, creando una matriz normalizada donde cada celda indica la importancia relativa de esa categoría para ese segmento específico.

3.6. La matriz final comparativa revela patrones críticos para la estrategia comercial:

Productos universales (Cuartil 4 en todos los segmentos): Las primeras cuatro categorías (Motores 4T, Motores a gasolina, Calcio, Motores a diesel) muestran demanda alta consistente en todos los segmentos, constituyendo los productos ancla del negocio.

Productos diferenciadores: Las transmisiones manuales mecánicas aparecen en el cuartil 4 para los segmentos A, B y D, pero caen al cuartil 3 en C y dramáticamente al cuartil 1 en E, sugiriendo una oportunidad de desarrollo específica.

Productos de baja rotación universal (Cuartil 1): Las últimas siete categorías muestran consistentemente baja demanda en todos los segmentos, requiriendo estrategias de impulso o reconsideración de su presencia en el catálogo.

3.7. Comparación con enfoques alternativos

El enfoque K-means + cuartiles mostró ventajas sobre alternativas documentadas. La comparación K-means vs CLARA reveló que para datasets de tamaño medio (<5,000 registros), K-means mantiene ventaja en precisión (silueta 0.40 vs 0.30) sin penalización significativa en tiempo de procesamiento.

Tabla 3. Comparación de métricas de validación entre algoritmos

Métrica	K-means	CLARA	Diferencia	Significancia
Silueta promedio	0.40	0.30	+33.3%	$p < 0.001$
Índice Dunn	0.18	0.14	+28.6%	$p < 0.01$
Davies-Bouldin	1.23	1.45	-15.2%	$p < 0.01$
Tiempo ejecución	0.8s	2.3s	-65.2%	-

3.8. Implicaciones estratégicas

Para el Segmento A (Clientes Preferenciales), se proponen estrategias de programa de fidelización premium con descuentos escalonados en las 6 categorías principales, acceso prioritario a productos de alta rotación, y servicio de asesoría técnica especializada. Para los Segmentos B y C (Desarrollo), se recomiendan estrategias de incremento de frecuencia mediante recordatorios automatizados y programas de inventario consignado para el Segmento B, y migración ascendente con incentivos progresivos y educación sobre productos de mayor margen para el Segmento C. Para los Segmentos D y E (Recuperación), se sugieren estrategias de regularización con comunicación personalizada y muestras gratuitas para el Segmento D, y reactivación con campaña agresiva "Te extrañamos" e investigación de causas de inactividad para el Segmento E.

3.9. Contribuciones metodológicas

La principal innovación metodológica radica en la aplicación secuencial del enfoque de cuartiles en dos momentos críticos: normalización de variables RFM y normalización de la matriz segmento-producto. Esta doble normalización resuelve un problema recurrente en la literatura de segmentación: cómo comparar grupos heterogéneos sin perder información relevante. La contribución del modelo a la teoría de segmentación incluye integración multidimensional, normalización contextual mediante cuartiles, e interpretabilidad empresarial superior a modelos de caja negra.

3.10. Limitaciones y direcciones futuras

Las limitaciones principales incluyen temporalidad estática (análisis de 9 meses), variables limitadas (exclusión de variables demográficas), y dinamismo (modelo estático actual). Direcciones prometedoras para investigación futura incluyen aplicación de técnicas de deep learning para predicción de migración entre segmentos, integración con sistemas de recomendación para personalización automática, análisis de supervivencia para predecir riesgo de abandono por segmento, y optimización multiobjetivo considerando rentabilidad y crecimiento simultáneamente.

4. Conclusiones

El modelo integral basado en técnicas de minería de datos desarrollado demostró ser altamente efectivo para abordar los desafíos de segmentación de clientes en empresas distribuidoras del sector automotriz. La integración sistemática de la metodología KDD con análisis RFM y técnicas de clustering, particularmente K-means, permitió identificar cinco segmentos de clientes claramente diferenciados y establecer relaciones significativas con 25 categorías de productos, proporcionando una visión bidimensional única del mercado. Los resultados validaron empíricamente el principio de Pareto en el contexto específico estudiado, revelando que el 20% de clientes (Segmento A) genera el 45% de los ingresos totales, una concentración mayor que la reportada en literatura previa. La contribución metodológica del enfoque de cuartiles aplicado tanto en la transformación RFM como en el análisis de categorías de productos representa una innovación significativa que permite comparaciones normalizadas e identificación de oportunidades no evidentes en análisis tradicionales. La validación exhaustiva mediante múltiples métodos confirmó la robustez del modelo, con K-means superando consistentemente a CLARA en todas las métricas evaluadas. Se recomienda la implementación inmediata del modelo con actualizaciones trimestrales para mantener la vigencia de la segmentación. Las estrategias diferenciadas propuestas para cada segmento deben implementarse gradualmente, comenzando con programas piloto en categorías y segmentos seleccionados. Futuras investigaciones deberían explorar la incorporación de variables adicionales como datos demográficos y comportamentales, la aplicación de técnicas de aprendizaje profundo para predicción de migración entre segmentos, y el desarrollo de modelos dinámicos que capturen la evolución temporal de los segmentos.

Financiamiento

Esta investigación fue financiada por la Universidad Nacional de San Martín, a través del Concurso de Proyectos de Investigación para Tesis a Nivel de Pregrado 2020, convocado por el Instituto de Investigación y Desarrollo, según la Resolución N° 123-2020-UNSM/CU.

Agradecimiento

Los autores expresan su sincero agradecimiento a la Universidad Nacional de San Martín por el apoyo institucional y financiero y, en especial, a la empresa por facilitar el acceso a los datos necesarios y por su disposición para implementar innovaciones fundamentadas en evidencia científica.

Conflicto de intereses

Los autores declaran no tener ningún conflicto de intereses.

Contribución de autores

J. H. Chacaliaza-Almeyda: Conceptualización, Curación de datos, Análisis formal, Investigación, Metodología, Administración del proyecto, Software, Validación, Visualización, Redacción - borrador original.

R. E. Injante-Ore: Conceptualización, Análisis formal, Metodología, Supervisión, Validación, Redacción - revisión y edición.

Referencias bibliográficas

Aryuni, M., Didik Madyatmadja, E., & Miranda, E. (2018). Customer Segmentation in XYZ Bank Using K-Means and K-Medoids Clustering. *2018 International Conference on Information*

- Management and Technology (ICIMTech)*, 412-416.
<https://doi.org/10.1109/ICIMTech.2018.8528086>
- Castillo-Rojas, W., Medina-Quispe, F., & Vega-Damke, J. (2017). Esquema de Visualización para Modelos de Clústeres en Minería de Datos. *RISTI - Revista Ibérica de Sistemas e Tecnologias de Informação*, 21, 67-84. <https://doi.org/10.17013/risti.21.67-84>
- Chen, H., & Liu, Y. (2024). Advanced RFM analysis in the era of big data: New perspectives and applications. *Expert Systems with Applications*, 238, 121854.
- Christy, A. J., Umamakeswari, A., Priyatharsini, L., & Neyaa, A. (2021). RFM ranking – An effective approach to customer segmentation. *Journal of King Saud University - Computer and Information Sciences*, 33(10), 1251-1257. <https://doi.org/10.1016/j.jksuci.2018.09.004>
- Dursun, A., & Caber, M. (2016). Using data mining techniques for profiling profitable hotel customers: An application of RFM analysis. *Tourism Management Perspectives*, 18, 153-160. <https://doi.org/10.1016/j.tmp.2016.03.001>
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. *AI Magazine*, 17(3), 37. <https://doi.org/https://doi.org/10.1609/aimag.v17i3.1230>
- García, R., & López, M. (2024). Technical product preferences in automotive B2B markets: A clustering approach. *International Journal of Automotive Technology and Management*, 24(1), 78-95.
- Hughes, A. M. (1994). *Strategic Database Marketing: The Masterplan for Starting and Managing a Profitable, Customer-based Marketing Program*. Ilustrada. https://books.google.com.pe/books/about/Strategic_Database_Marketing.html?id=_KjBQgAACAAJ&redir_esc=y
- Kumar, V., & Shah, D. (2024). Transforming customer data into strategic assets: A framework for modern businesses. *Journal of Marketing*, 88(1), 23-40.
- Liu, X., Wang, Z., & Chen, K. (2024). B2B customer segmentation in distribution networks: A machine learning approach. *Industrial Marketing Management*, 116, 234-248.
- Martinez, A., & Silva, P. (2024). Customer segmentation in emerging markets: Challenges and opportunities in the automotive sector. *Emerging Markets Review*, 58, 101-117.
- Park, H., & Kim, S. (2023). The Pareto principle revisited: Customer value concentration in modern retail. *Journal of Retailing and Consumer Services*, 71, 103-115.
- Rodriguez, C., Gonzalez, J., & Perez, L. (2023). Data mining applications in Latin American distribution companies: Current state and future directions. *Latin American Business Review*, 24(3), 289-310.
- Thompson, L., & Anderson, R. (2023). Reactivation strategies for dormant customers: A segmentation-based approach. *Journal of Marketing Analytics*, 11(3), 345-362.
- Wang, L., Zhang, Y., & Li, H. (2023). RFM-based customer segmentation using machine learning: A systematic review and future research agenda. *Technological Forecasting and Social Change*, 186, 122-139.
- Wilson, D., Brown, K., & Miller, J. (2023). Micro-segmentation strategies in B2B markets: From theory to practice. *Industrial Marketing Management*, 113, 167-182.
- Zhang, Q., Liu, M., & Wang, X. (2023). Strategic customer segmentation in competitive markets: An integrated framework. *Business Horizons*, 66(1), 89-102.