



## EL USO DE LA ANALÍTICA DE DATOS PARA LA SEGMENTACIÓN Y PREDICCIÓN DEL COMPORTAMIENTO DE LOS CLIENTES

Davinson Mosquera González

Ingeniero Administrador / Especialista en Analítica / Magíster en Ingeniería-Analítica

[davmosqueragon@unal.edu.co](mailto:davmosqueragon@unal.edu.co)

Tesis de referencia



### Problema

En las últimas décadas, las empresas han experimentado un entorno cada vez más competitivo, donde solo aquellas que aporten verdadero valor a sus clientes, con base en relaciones sólidas y rentables en el tiempo, sobrevivirán. Es por esto por lo que las organizaciones están en una constante búsqueda para identificar, atraer y retener a sus clientes (Tsai *et al.*, 2019), lo que se conoce como Customer Relationship Management (CRM), y que trae consigo el concepto de segmentación de clientes, el cual consiste en el agrupamiento de los diferentes clientes de una organización en subgrupos más pequeños que el total, donde se tiene como objetivo que clientes con características similares (sociodemográficas o transaccionales) se encuentren en el mismo grupo y a partir de allí se puedan hacer estrategias segmentadas o dirigidas que permitan la optimización de recursos (Chatterjee *et al.*, 2021).

En este contexto, el uso de los datos del comportamiento de compra y consumo de los clientes, frecuentemente denominado “datos transaccionales”, representa un insumo fundamental para crear estrategias que permitan mejorar el relacionamiento con los clientes y, en consecuencia, segmentarlos. Para este fin, la analítica de los datos se hace esencial; esta puede ser entendida como un proceso en el cual se utilizan técnicas y herramientas para generar información, patrones, tendencias y,

en general, conocimiento a partir de los datos, contribuyendo así con la toma de decisiones en las organizaciones.

Históricamente, una de las técnicas más utilizada para la segmentación de clientes es el modelo Recencia, Frecuencia, Monto, del inglés Recency, Frequency, Monetary (RFM), en el que a partir de datos transaccionales de los clientes se crean grupos de clientes con comportamientos de compra similares (Heldt *et al.*, 2021; Rathi y Ravi, 2017); sin embargo, el uso de esta técnica presenta diferentes limitaciones, tales como: 1) se basa en momentos puntuales del tiempo y 2) restringen su capacidad de segmentación al comportamiento pasado de los clientes, ignorando el pronóstico del comportamiento futuro.

En consecuencia, la pregunta problematizadora del presente artículo es ¿cómo realizar los procesos de segmentación de clientes a partir de la analítica de datos y cómo incorporar la predicción de su futuro comportamiento? Teniendo en cuenta que este comportamiento futuro es entendido como el valor monetario del cliente, y que representa el aporte económico que el cliente realiza a la empresa durante el periodo de relación con esta; en la literatura formal es llamado *Customer Lifetime Value*. Para ilustrar lo anterior, la figura 6.1 representa un caso hipotético de clientes de una empresa; cada círculo o punto

equivale a una compra en el tiempo. El reto radica en utilizar ese comportamiento pasado para realizar estimaciones o predicciones de comportamientos futuros y, finalmente, usar esta predicción para establecer segmentaciones de clientes.

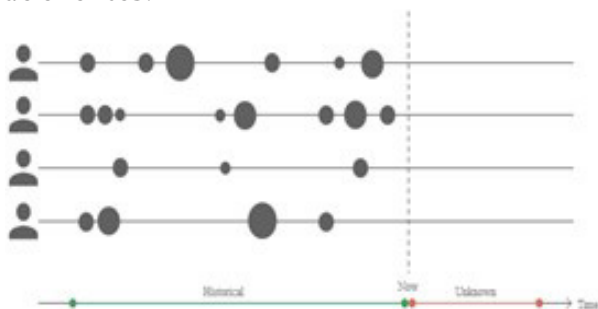


Figura 6.1 Representación histórica de compras de clientes  
Fuente: Google Architecture Center (2021).

### Abordaje metodológico

Para el abordaje de la pregunta problematizadora se propone un enfoque metodológico mixto, en el que en un primer momento se realiza una revisión de la literatura, soportada en el uso de técnicas bibliométricas, en el cual se tiene como objetivo indagar respecto a la evolución del campo de conocimiento, así como identificar las tendencias y las técnicas emergentes que puedan ser usadas en la solución de la problemática. En un segundo momento, se realiza una agrupación por las tendencias encontradas y resultados de un caso de aplicación real.

Al detallar respecto a la metodología propuesta, el análisis bibliométrico es una técnica que utiliza métodos matemáticos y estadísticos para medir la cantidad y calidad de la producción científica de un campo del conocimiento (Cadavid *et al.*, 2012). En este análisis, se definió una ecuación de búsqueda enfocada en indagar respecto a dos preguntas: ¿Cuáles son las técnicas y los modelos usados para predecir el valor monetario del cliente y lograr su segmentación? ¿Cuáles son las tendencias en la medición y predicción del valor monetario del cliente para su segmentación?

### Resultados

Siguiendo con la metodología propuesta, en la búsqueda realizada en Scopus se encontraron 499 documentos científicos, comprendidos entre el periodo 1987 y 2021, tal como se presenta en la figura 6.2. En este histórico se observa que el campo de conocimiento ha estado en crecimiento, en especial desde el año 2003, cuando se experimentó un incremento exponencial seguido de la presencia de diferentes ciclos, que pueden ser explicados por las temáticas emergentes que han marcado cada uno de ellos.

Al realizar el análisis de las publicaciones encontradas, se nota que el campo de conocimiento evolucionó a lo largo de los años; así, entre 1987 y el 2002 se hizo extensivo el uso de técnicas paramétricas y estadísticas para estimar el valor monetario de los clientes como insumo para su segmentación, como es el caso del modelo Pareto/NBD. Al 2013 se propusieron nuevas técnicas paramétricas o modificaciones a las existentes. Finalmente, del 2014 al 2021 se han presentado nuevos enfoques que usan la analítica de datos y el aprendizaje de máquinas para la predicción del valor monetario de los clientes en la segmentación de clientes, como son Random Forest y Support Vector Machine.

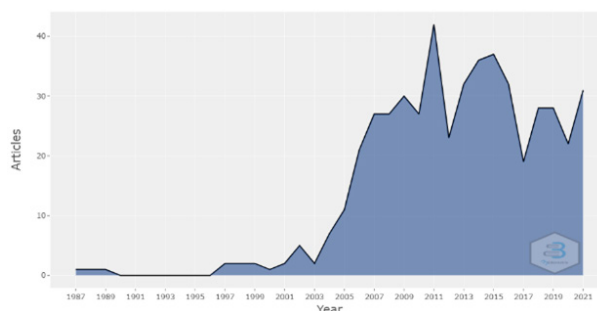


Figura 6.2 Producción científica histórica del campo de conocimiento  
Fuente: elaboración propia.

En la tesis titulada *Método para la segmentación de clientes incorporando la predicción del valor monetario del cliente como una variable de segmentación* (2022) se presenta un análisis experimental que compara los resultados de dos procesos de segmentación de clientes, en

el que uno de ellos incluye la predicción del valor monetario del cliente como una variable, mientras que el otro no la incorpora. Este análisis se realizó tomando como insumo la base de datos abierta de la “Online Retail II Data Set”, correspondiente a una empresa minorista en línea del Reino Unido, que vende artículos de regalo para toda ocasión. Fue fundada en 1981, su canal de venta principal es Amazon y dispone de ochenta empleados. Este conjunto de datos contiene todas las transacciones o compras que se produjeron entre los años 2009 y el 2011, con un total de 1 067 371 registros transaccionales y un total de 5833 clientes (Chen, 2019).

Los resultados muestran que, al incorporar la variable de predicción del valor monetario del cliente, se encuentran segmentos de clientes, para la empresa, que agregan un 12 % de valor económico mayor respecto al otro grupo que no incorporó la predicción del valor monetario del cliente (Mosquera-González, 2022); los resultados se presentan en la figura 6.3.

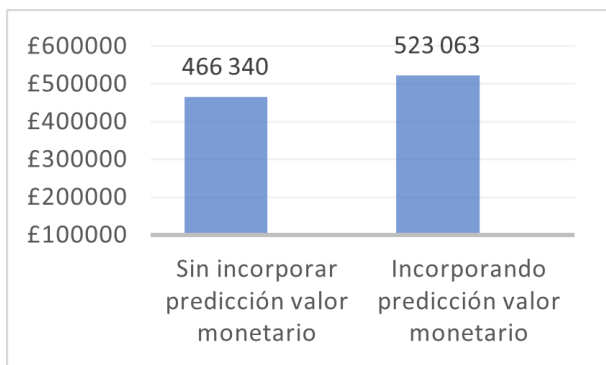


Figura 6.3 Comparativo de valor monetario total a partir de segmentos de clientes identificados  
Fuente: elaboración propia.

## Conclusiones

A partir de los hallazgos presentados se concluye que existe evidencia a favor del uso de la analítica de datos en los procesos de segmentación de clientes, principalmente respecto al uso de técnicas de aprendizaje de máquinas para realizar la predicción del valor monetario del cliente. Este enfoque sugiere que, de esta manera se identifican los segmentos

de clientes más valiosos que los reconocidos utilizando técnicas tradicionales basadas en el comportamiento pasado de los clientes. Para finalizar, se recomienda que, para la aplicación de métodos similares en la industria real, las organizaciones implementen estrategias de CRM, lo que les permitirá disponer de los datos transaccionales del comportamiento de compra de sus clientes.

## Referencias

Cadavid, L., Awad, G. y Franco, C. (2012). Análisis bibliométrico del campo modelado de difusión de innovaciones. *Estudios Gerenciales*, 28, 213-236. <http://www.scielo.org.co/pdf/eg/v28nspe/v28nspea12.pdf>.

Chatterjee, S., Rana, N. P., Tamilmani, K. y Sharma, A. (2021). The effect of AI-based CRM on organization performance and competitive advantage: An empirical analysis in the B2B context. *Industrial Marketing Management*, 97, 205-219. <https://doi.org/10.1016/j.indmarman.2021.07.013>.

Chen, D. (2019). *Online Retail II Data Set*. <https://archive.ics.uci.edu/ml/datasets/Online+Retail+II>.

Google Architecture Center (2021). <https://cloud.google.com/architecture?hl=es-419>.

Heldt, R., Silveira, C. S. y Luce, F. B. (2021). Predicting customer value per product: From RFM to RFM/P. *Journal of Business Research*, 127, 444-453. <https://doi.org/10.1016/j.jbusres.2019.05.001>.

Mosquera-González, D. (2022). *Método para la segmentación de clientes incorporando la predicción del valor monetario del cliente como una variable de segmentación*. [Tesis de maestría]. Universidad Nacional de Colombia, Medellín, Colombia. <https://repositorio.unal.edu.co/handle/unal/81631>.

Rathi, T. y Ravi, V. (2017). Customer lifetime value measurement using machine learning techniques. En *Artificial Intelligence: Concepts, methodologies, tools, and applications* (pp. 3013-3022). <https://doi.org/10.4018/978-1-5225-1759-7.ch124>.

Tsai, T., Lin, C. y Prasad, M. (2019). An Intelligent customer churn prediction and response framework. *2019 IEEE 14th International Conference on Intelligent Systems and Knowledge Engineering (ISKE)*, 928-935. <https://doi.org/10.1109/ISKE47853.2019.9170380>.