



UNIVERSIDAD
PRIVADA
DEL NORTE

FACULTAD DE NEGOCIOS

CARRERA DE ADMINISTRACIÓN Y NEGOCIOS INTERNACIONALES.

“TENDENCIAS DEL COMPORTAMIENTO DE CONSUMIDORES MEDIANTE HERRAMIENTAS DE DATA MINING, EN SUPERMERCADOS DEL PERÚ: UNA REVISIÓN DE LA LITERATURA CIENTÍFICA”.

Trabajo de investigación para optar al grado de:

Bachiller en ADMINISTRACIÓN Y NEGOCIOS INTERNACIONALES.

Autores:

Libardo Enrique Álvarez Montes

Carla Luciana Hurtado Cavero

Asesor:

Mg. Luis Clemente Baquedano Cabrera

Trujillo - Perú

2020

DEDICATORIA

Libardo:

A mis padres, quienes me han brindado su apoyo, amor y consejos. A mi hermano, que estuvo acompañándome todas las noches en el proceso para hacer esto posible.

Carla:

A Carmen, por dedicarme muchos años de su vida brindándome cariño incondicional, apoyo y confianza. Hoy no estás, pero yo sigo aquí gracias a ti.

AGRADECIMIENTO

Libardo:

Agradezco a mis padres porque sin ellos no sería la persona que soy. A mis amigos, que, a pesar de la distancia, siempre cuento con ellos. A mi profesor de Taller de Tesis Luis Baquedano por aportar su tiempo y conocimientos. A Carla por brindarme su amistad desde que inició esta etapa universitaria.

Carla:

Quiero agradecer a mis padres por su esfuerzo que permite mi formación profesional, a mi asesor Luis Baquedano por su apoyo en el desarrollo de este estudio, a mis docentes universitarios por facilitarme conocimientos y enseñanzas, a Libardo y amigos de la universidad y de la vida por brindarme su amistad y soporte para cumplir mis metas.

Tabla de contenido

	Página
DEDICATORIA	2
AGRADECIMIENTO	3
ÍNDICE DE TABLAS	5
ÍNDICE DE FIGURAS	6
RESUMEN	7
CAPÍTULO I. INTRODUCCIÓN	8
CAPÍTULO II. METODOLOGÍA	12
CAPÍTULO III. RESULTADOS	14
CAPÍTULO IV. DISCUSIONES Y CONCLUSIONES	36
REFERENCIAS	38

ÍNDICE DE TABLAS

Tabla 1 Matriz de artículos	17
Tabla 2 Matriz de indagación	20
Tabla 3 Características de los estudios según el tipo de documento	27
Tabla 4 Características de los estudios según año de publicación	27
Tabla 5 Características de los estudios según Revista de publicación	28

ÍNDICE DE FIGURAS

Figura 1 Proceso de selección de artículos.	15
Figura 2 Proceso de creación de reglas de asociación usando el algoritmo FSA-Red, propuesto por Sulianta et al., 2013.	30

RESUMEN

A pesar de que el Data Mining es una herramienta clave para comprender el comportamiento de clientes de tiendas minoristas como supermercados, hace falta una revisión exhaustiva de la literatura que ayude a comprender mucho mejor su aplicación y los beneficios que esta trae. Esta revisión proporciona una base de datos dentro de un periodo de diez años, entre el 2009 y 2019, de donde en un primer momento se extrajeron 48 artículos por tener relación con la aplicación de Data Mining para encontrar tendencias de comportamiento de consumidores, se extrajeron de nueve bases de datos especializadas: Emerald Insight, ResearchGate, IEEE Xplore, ScienceDirect, Semantic Scholar, Google Scholar, Cuaderno Activa, Taylor and Francis y Mary Ann Liebert, INC. Los resultados demuestran que para entender el comportamiento del consumidor se debe minar: i) Su ruta de compras, temporadas, cantidades y preferencias con el fin de organizar los productos en la tienda y realizar promociones acorde a su perfil; ii) El CLV con el fin de separarlos en segmentos y enfocarse en ellos con mayor claridad; iii) Sus necesidades y quejas para mejorar su CRM; iv) Sus percepciones y requisitos para aplicar estrategias de marketing más precisas; v) Qué productos buscan y comparan en tiendas de retail online, tema del cual se tiene poca información, por lo que se recomienda ahondar a detalle para otras investigaciones. Todo ello mediante diferentes técnicas de DM y se demostró que las redes neuronales, seguidas por los árboles de decisión y las técnicas de reglas de asociación son recomendadas para los que no son expertos en DM. Nuestra investigación sirve como un derrotero para la futura creación de conocimiento sobre los temas abordados.

PALABRAS CLAVES: Data mining, comportamiento del consumidor, Valor de Vida del Cliente, Gestión de Relaciones con Clientes, Tiendas Minoristas, Diseño de tienda.

CAPÍTULO I. INTRODUCCIÓN

En la actualidad, el entorno donde se desarrollan las tiendas minoristas está cambiando gracias al avance de tecnologías que afectan la búsqueda, selección y comparación de productos por parte de los clientes, mejorando así su experiencia de compra (Pantano y Naccarato, 2010, citado por Alfian *et al.*, 2018). Es por esto que, las empresas de este rubro que buscan obtener una ventaja competitiva utilizando información y conocimiento relacionados a su empresa, deben tener un fuerte enfoque en estas y tener claro el vínculo de estas a su organización para tomar decisiones importantes (Ghazanfari *et al.*, 2011, citado por Altuntas, 2017; Davenport, 2010); sin embargo, muchas veces la incapacidad de descubrir información valiosa oculta en los datos impide a estas organizaciones transformen estos datos en conocimiento útil y sobretodo, valioso para ellos mismos (Berson *et al.* 2000, citado por Ngai *et al.*, 2009). Para lograr esto con éxito, las empresas deben saber cómo se toman y ejecutan las decisiones, cómo es que estas se pueden mejorar y cómo se usa la información para respaldar las mismas (Davenport, 2010). Un claro ejemplo del uso de estas herramientas es cuando se estudia el comportamiento del cliente viendo la atracción hacia ciertos artículos y la cantidad promedio comprada, con esta información, los minoristas pueden tomar decisiones respecto al diseño de la tienda, considerando perspectivas como precio, cantidad y beneficio; encontrando así asociaciones entre productos, lo cual le permite al comercio minorista ubicar otros productos cercanos en función a estas asociaciones. (Altuntas, S. 2017).

Un punto importante a considerar para evaluar el rendimiento de un negocio es el análisis del comportamiento del cliente, el cual se realiza a través de la comprensión de

clientes pasado y actuales para identificar quienes son los posibles futuros clientes y su comportamiento (Mudimigh *et al.*, 2009); de esta manera, se utilizan todos los datos existentes de los clientes para construir un modelo de comportamiento del cliente con la finalidad de predecir el comportamiento del cliente potencial (Hsieh y Chu, 2009). Analizar cuidadosamente el comportamiento del consumidor es importante para discriminar mejor y asignar recursos de manera más efectiva al grupo de clientes más rentable que cumplan ciertos criterios de rentabilidad en función de sus necesidades individuales o comportamientos de compra, en lugar de dirigirse u ofrecer incentivos a todos los clientes por igual (Golmah y Mirhashemi, 2012). Para fortalecer el comportamiento del cliente día a día, la mayoría de empresas eligen la satisfacción del cliente como su principal indicador de desempeño (Mudimigh *et al.*, 2009) y a su vez, los clientes evalúan la brecha entre sus expectativas y su experiencia antes y después de su consumo para definir su satisfacción (Hofacker *et al.*, 2016). Después de analizar este punto, recién una empresa podrá revisar cuáles son sus atributos, compararlos con los requisitos de sus clientes y determinar si puede ser atractivo para futuros clientes (Mudimigh *et al.*, 2009).

En el ámbito empresarial, la minería de datos puede ser utilizada como una herramienta de inteligencia comercial para ayudar a estas organizaciones a descubrir conocimientos escondidos en una enorme cantidad de datos (Wang y Wang, 2008, citado por Altuntas, 2017; Ngai *et al.*, 2009). La minería de datos se puede definir como un proceso que usa técnicas estadísticas, matemáticas, de inteligencia artificial y de aprendizaje automático que permite extraer información útil de una base de datos grande y posteriormente, obtener conocimientos (Turban *et al.*, 2007, citado por Ngai *et al.*, 2009). Por lo tanto, puede ser una herramienta muy útil para la segmentación de mercado, perfil del

cliente, análisis de riesgos, etc. o incluso, para decisiones aplicables a estrategias de marketing (Hsieh y Chu, 2009) debido a que es un proceso de descubrimiento de entendimientos interesantes como patrones, asociaciones, cambios, anomalías encontrados en grandes repositorios de información (Hui y Jha, 2000; Chua y Lan, 2005, citado por Huang y Hsueh, 2010).

Entonces, podemos decir que el análisis del comportamiento del consumidor se realiza a través de la comprensión de clientes pasado y actuales para identificar quienes son los posibles futuros clientes y su comportamiento (Mudimigh et al., 2009) y es importante para discriminar mejor y asignar recursos de manera más efectiva al grupo de clientes más rentable, en lugar de dirigirse u ofrecer incentivos a todos los clientes por igual (Golmah y Mirhashemi, 2012). Por otro lado, la minería de datos puede ser utilizada como una herramienta de inteligencia comercial para ayudar a estas organizaciones a descubrir conocimientos escondidos en una enorme cantidad de datos (Wang and Wang, 2008; Ngai et al., 2009). Por ello, se considera que utilizar herramientas de minado de datos será de gran ayuda para identificar patrones, cambios, asociaciones, etc. en la conducta de los clientes, escondidos en una big data, que puedan ser utilizados en la toma de decisiones de gerentes en sus organizaciones.

En este estudio se busca analizar el comportamiento de los consumidores en el sector retail de supermercados usando de herramientas de minado de datos para poder responder ¿cuáles son las tendencias del comportamiento de consumidores mediante herramientas de data mining en supermercados, entre los años 2009 y 2019?

El objetivo principal de este trabajo de investigación es ayudar a identificar, clasificar, analizar toda la información existente acerca de las tendencias de comportamiento de consumidor del sector retail, para proporcionar a los investigadores una visión general de los estudios dentro de este campo, cuáles son los conceptos que se han explorado, cuáles son los principales hallazgos a través de las herramientas de data mining y qué métodos de investigación se utilizan en la mayoría de investigaciones.

CAPÍTULO II. METODOLOGÍA

Para el desarrollo de esta investigación se planea recolectar diferentes fuentes de información las cuales deben considerar las variables que se quieren abordar como son el comportamiento del consumidor y el minado de datos. Asimismo, se consideró que las fuentes de información deben ser entre los años 2009 y 2019 porque se consideró un límite de diez años de antigüedad. El idioma de las fuentes puede ser tanto inglés como español por el buen manejo de ambos idiomas. Estos estudios deben haber sido publicados para investigaciones de estudios universitarios y, por último, de preferencia que tengan el formato IMRD.

Las bibliotecas virtuales consultadas fueron Emerald Insight, que incluye revistas y libros relevantes de diversas áreas como contabilidad y finanzas, gestión, marketing, ingeniería, etc.; ScienceDirect, que es un sitio web que proporciona acceso por suscripción a una gran base de datos de investigación científica, técnica y médica; además, se consultó IEEE Xplore, que es una base de datos de investigaciones confiables como revistas, conferencias, libros electrónicos, etc.; a su vez, se utilizó ResearchGate que es un sitio europeo de redes sociales comerciales para que científicos e investigadores compartan documentos, resuelvan dudas, etc.; asimismo, se consultó Semantic Scholar que utiliza Inteligencia Artificial para comprender la semántica de la literatura científica para ayudar a los académicos a descubrir investigaciones relevantes; entre otras bases de datos reconocidas como ProQuest, Google Scholar, Cuaderno Activa, Mary Ann Liebert, INC y Taylor and Francis.

Una vez definidas las bibliotecas de búsqueda, se procedió a buscar palabras clave en ambos idiomas (comportamiento del consumidor, customer behavior, consumer behavior, minería de datos, minado de datos, data mining, retail, big data, etc.), a su vez, fue necesario especificar los tipos de documentos que se querían obtener, en este caso, investigaciones originales, artículos científicos, tesis, artículos de revisión y que contengan métodos o metodologías. El periodo fue un factor notable, ya que se buscaban documentos que estuvieran entre el rango de tiempo delimitado, en este caso, los últimos diez años; de esta manera se encontraron un total de 48 artículos.

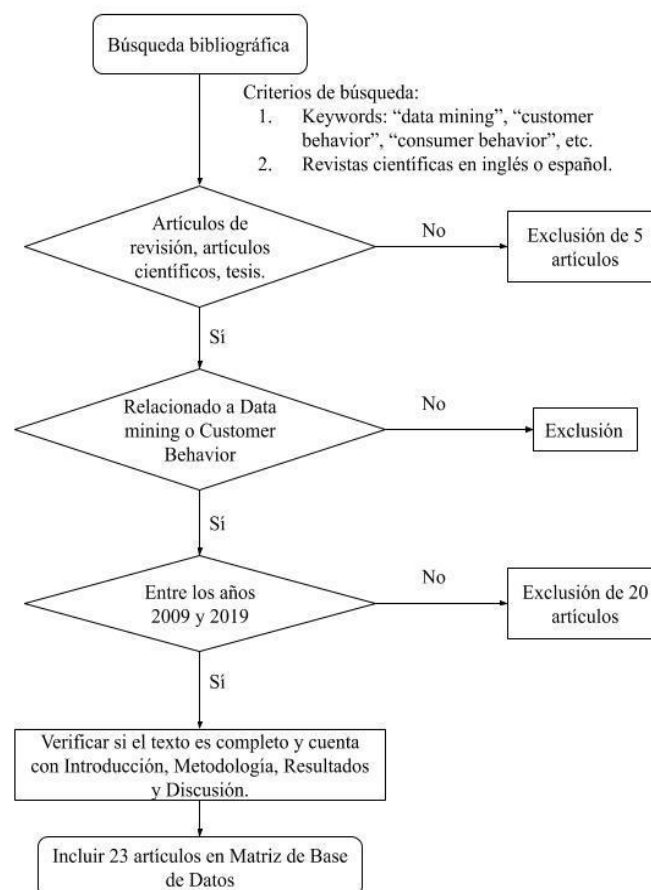
Se descartaron un total de 25 fuentes de información porque a pesar de considerar las variables requeridas y cumplir con otros requisitos, no estaban dentro del rango de tiempo límite de diez años que se consideró para esta investigación ni eran artículos científicos. Por otro lado, se incluyó un total de 23 fuentes que sí cumplieron con los criterios de inclusión.

Se tuvo en cuenta el año de publicación, su país de procedencia y la institución que desarrolló la investigación, los cuales están ubicados muchas veces en la primera página del artículo, así como también palabras clave a la hora de recolectar información que se encontraron la mayoría de veces después del resumen o abstract de la fuente de información o en la revista de donde se extrajo. Una vez seleccionadas, se procedió a verificar los procesos y métodos expuestos para verificar su tipo de estudio y finalmente incluir nueve artículos en nuestra investigación.

CAPÍTULO III. RESULTADOS

Cada artículo considerado debió cumplir con los criterios de inclusión para lo cual, la selección de artículos se evaluó considerando el proceso mostrado en Fig. 1. El proceso de selección de artículos se repitió con cada revista revisada, usando las palabras clave relacionadas a nuestra investigación como “data mining”, “customer behavior”, etc. y considerando los artículos en inglés y español. Se verificó que sean del tipo de artículo considerado criterios de inclusión, relacionados a nuestro problema de investigación y que cuenten con Introducción, Metodología, Resultados y Discusión.

Figura 1: Proceso de selección de artículos.



Fuente de información: Autoría propia

De acuerdo a la revisión sistemática realizada, teniendo como apoyo bases de datos especializadas como Emerald Insight, Google Scholar, Semantic Scholar, Science Direct, World Scientific, IEEE Xplore, etc., se logró obtener un total 48 artículos; de los cuales se seleccionaron 23. Llegando a la conclusión de incluir estos 23 artículos como se muestra en Tabla 1, donde se muestran los artículos detallados según la revista donde se encontraron, sus autores, universidad, año de publicación y tipo de investigación; de la misma manera, en Tabla 2, se detalló la indagación de los estudios, ordenando así los objetivos, resultados y conclusiones de ellos.

Tabla 1: Matriz de artículos

No.	Base de datos	Título	Universidad	Autor	Año	Tipo de Investigación		
						Tesis	AC	AR
1	Emerald Insight	Customer behavior analysis using real-time data processing: A case study of digital signage-based online stores.	Dongguk University	Alfian, G., <i>et al.</i>	2018		x	
2	Emerald Insight	A novel approach based on utility mining for store layout: a case study in a supermarket.	Yildis Technical University	Altuntas, S.	2017		x	
3	Google Scholar	A study of discovering customer value for CRM: Integrating customer lifetime value analysis and data mining techniques.	National Chiao Tung University	Chen, C. W., <i>et al.</i>	2009		x	
4	ProQuest	Data mining for the online retail industry: A case study of RFM model-based customer segmentation using data mining.	London South Bank University	Chen, D., <i>et al.</i>	2012		x	
5	Semantic Scholar	Business Intelligence and Organizational Decisions.	Babson College	Davenport, T.H.	2010		x	
6	Science Direct	Automating the Extraction of Static Content and Dynamic Behaviour from e-Commerce Websites.	University of Porto	Dias, J. y Ferreira, H.	2017		x	

7	Science Direct	Unveiling the relationship between the transaction timing, spending and dropout behavior of customers.	ESSEC School, University, Katholieke Universiteit Leuven	Business Tilburg	Glady, N., et al.	2014	x	
8	Google Scholar	Implementing A Data Mining Solution To Customer Segmentation For Decayable Products - A Case Study For A Textile Firm.	Islamic University	Azad	Golmah, V. y Mirhashemi, G.	2012	x	
9	Google Scholar	An Empirical Assessment of Customer Lifetime Value Models within Data Mining.	Bakircay University, Dumlupinar University	Sakarya University	Hiziroglu, A., et al.	2018	x	
10	Emerald Insight	Big Data and consumer behavior: imminent opportunities.	La Trobe University		Hofacker, C. F., et al.	2016		x
11	Google Scholar	Enhancing Consumer Behavior Analysis by Data Mining Techniques.	Tamkang University		Hsieh, N. C., y Chu, K. C.	2009	x	
12	Taylor and Francis	Customer behavior and decision making in the refurbishment industry - A data mining approach.	National Kaohsiung University of Applied Sciences, Tung Fang Institute of Technology		Huang, C. F. y Hsueh, S. L.	2010	x	
13	Science Direct	Big data analysis to Features Opinions Extraction of customer.	Université de Tunis		Mars, A. y Gouider, M.	2017	x	
14	IEEE Xplore	Efficient implementation of data mining: Improve customer's behaviour.	King Saud University		Mudimigh, A., et al.	2009	x	
15	Science Direct	Application of data mining techniques in customer relationship management: A literature review and classification.	The Hong Kong Polytechnic University, Tsinghua University		Ngai, E.W.T., et al.	2009		x
16	Cuaderno o Activa	Visualizaciones analíticas para la toma de decisiones en pequeñas y medianas empresas utilizando Data Mining	Tecnológico de Antioquia	de	Noreña, P., et al.	2016	x	
17	Mary Ann Liebert	Data Science and its Relationship to Big Data and Data-Driven Making.	New York University, Scientists, LLC	Data	Provost, F., Fawcett, T.	2013		x
18	Google Scholar	Data Mining Techniques: A Source for Consumer Behavior Analysis.	Shivaji University		Raorane, A. y Kulkarni, R.V.	2011		x

19	Research Gate	Mining utility-oriented association rules: An efficient approach based on profit and quantity.	Rayat and Bahra Institute of Engineering and Bio-Technology, RIMIT Institute of Engineering and Technology	Sandhu, P. S., et al.	2011	x
20	IEEE Xplore	Mining food industry's multidimensional data to produce association rules using apriori algorithm as a basis of business strategy.	Telkom Institute of Technology	Sulianta, F., et al.	2013	x
21	Taylor and Francis	A data mining approach to optimise shelf space allocation in consideration of customer purchase and moving behaviours.	Yuan-Ze University	Tsai, C.-Y. y Huang, S.-H.	2014	x
22	Research Gate	Mining of users access behavior for frequent sequential patterns from web logs.	Thiagarajar College of Engineering	Vijayalakshmi, S., et al.	2010	x
23	Google Scholar	Data Science, Predictive Analytics, and Big Data: A Revolution That Will Transform Supply Chain Design and Management.	University of Arkansas, Weber State University	Waller, M. y Fawcett, S.	2013	x

Tabla 2: Matriz de indagación

No.	Base de Datos	Título	Objetivos	Resultados	Conclusión
1	Emerald Insight	Customer behavior analysis using real-time data processing: A case study of digital signage-based online stores.	Proponer un análisis del comportamiento del cliente basado en el procesamiento de datos en tiempo real y la regla de asociación.	A medida que el número de clientes aumenta, el sistema propuesto fue capaz de procesar una cantidad gigantesca de información; entre más tiempo se dure la visita, más probable es que se compren productos.	Se ha concluido que existen hábitos de compra en los clientes, y de acuerdo a estos se pueden tomar decisiones para deleitar al cliente. DM es indispensable para la asociación de la información.
2	Emerald Insight	A novel approach based on utility mining for store layout: a case study in a supermarket.	El propósito de este artículo es proponer un enfoque novedoso basado en el minado de utilidades para el diseño de una tienda.	El diseño de tienda necesita un enfoque para mejorar la relación de ventas en el negocio minorista. Se propuso un enfoque para reorganizar su diseño en función de los hábitos de consumo de los clientes y mejorar su desempeño.	Los responsables de la toma de decisiones pueden utilizar el enfoque propuesto para garantizar la sostenibilidad del negocio con el apoyo de business intelligence.

- | | | | | | |
|---|------------------|---|---|---|---|
| 3 | Google Scholar | A study of discovering customer value for CRM: Integrating customer lifetime value analysis and data mining techniques. | El objetivo es desarrollar un proceso que integre el análisis CLV y el DM para mejorar la medición del valor del cliente y mejorar la eficiencia y efectividad del CRM. | Se muestra un grupo de clientes que pertenecen a “clientes de oro”, este grupo genera la mayor contribución a la empresa; Los gerentes deberían preocuparse por grupos de clientes que tienen poco potencial y baja lealtad con la empresa. | Se concluye que, los gerentes deberían prestar atención a ambos grupos, para de esta manera comprender las tendencias de compra para así saber qué estrategias de marketing utilizar. |
| 4 | ProQuest | Data mining for the online retail industry: A case study of RFM model-based customer segmentation using data mining. | El objetivo es ayudar a la empresa a comprender mejor a sus clientes; realizar un marketing más efectivo. | La organización puede comprender mejor a sus clientes explorando las asociaciones de productos a través de sus compras. | Se demuestra que hay dos pasos en el proceso de DM que son cruciales y requieren más tiempo: la preparación de datos y la interpretación y evaluación de modelos. |
| 5 | Semantic Scholar | Business Intelligence and Organizational Decisions. | Se describe un estudio de los intentos de las organizaciones para mejorar la toma de decisiones mediante el uso de información, etc. | Las empresas determinaron el siguiente proceso:
Paso 1: Enfoque estratégico en decisiones clave.
Paso 2: Suministro de información.
Paso 3: Diseño de decisión.
Paso 4: Ejecución de decisiones. | Las decisiones tienen impacto en todos los aspectos del desempeño organizacional, tanto en el dominio estratégico como en el táctico. |
| 6 | Science Direct | Automating the Extraction of Static Content and Dynamic Behaviour from e-Commerce Websites. | El objetivo es presentar un proceso capaz de extraer información útil sobre el comportamiento típico de los usuarios. | Se extrajo una muestra de usuarios y se aplicó el algoritmo propuesto, de los cuales se obtienen varios grupos de clientes atípicos, lo cual significa una gran tasa de rebote dentro de los clientes de este sitio web. | Este proceso se enriquece con el cruce de datos de diversas fuentes para encontrar relaciones y finalmente poder proponer un modelo de información para un sitio web. |
| 7 | Science Direct | Unveiling the relationship between the transaction timing, spending and dropout behavior of customers. | Se propone un modelo para estimar CLV que contabiliza el tiempo, gasto y decisiones de abandono de los clientes. | El enfoque tradicional para la valoración de CLV se centra en maximizar los flujos de efectivo a largo plazo de la empresa. De esta manera se puede pronosticar en función al CLV. | Se ofrece un método para mejorar las predicciones del modelo y generar nuevas ideas sobre el comportamiento de compra de los clientes. |

- | | | | | | |
|----|--------------------|--|---|--|--|
| 8 | Google Scholar | Implementing A Data Mining Solution To Customer Segmentation For Decayable Products - A Case Study For A Textile Firm. | Se desarrolla un modelo para agrupar a los clientes. Los simples modelos matemáticos para ello, conducen a ignorar los factores ambientales en el modelo. | El marketing uno a uno, los programas de fidelización y la gestión de quejas puede tener como objetivo mantener una relación a largo plazo con los clientes, de esta manera es más factible su identificación y su atracción. | Ayuda a interpretar de manera más fácil y precisa a los clientes objetivo. Se puede lograr obtener ganancias máximas con la situación de ganar-ganar para Albasco Co. |
| 9 | Google Scholar | An Empirical Assessment of Customer Lifetime Value Models within Data Mining. | El objetivo es comparar dos modelos diferentes de CLV dentro de la minería de datos. Se evaluó la segmentación de clientes con una base de datos de una empresa minorista. | Se utilizaron datos de marketing directo. Estos datos pertenecen a una organización sin fines de lucro que utiliza el correo directo para solicitar más contribuciones. | Se concluye que el uso de modelos CLV y técnicas de minería en conjunto brinda mucho potencial a las empresas para reconocer grupos de clientes de alto valor. |
| 10 | Emerald Insight | Big Data and consumer behavior: imminent opportunities. | Es evaluar cómo el estudio del comportamiento del consumidor puede beneficiarse de la presencia del Big Data. | Big Data tiene el potencial de ampliar la comprensión de cada etapa del proceso de toma de decisiones del consumidor. Ahora la naturaleza del ciclo de retroalimentación entre la teoría y los resultados puede cambiar bajo el peso del BD. | Se genera información relevante para todas las etapas del ciclo de toma de decisiones del consumidor. |
| 11 | Google Scholar | Enhancing Consumer Behavior Analysis by Data Mining Techniques. | Presenta un marco de análisis de comportamiento del consumidor en dos etapas. Identificar a los consumidores con este enfoque facilita el desarrollo de estrategias de marketing. | Después de revisar los perfiles de los clientes utilizando el inductor del árbol de decisión, los clientes con valores de cliente más altos podrían ser los grupos de clientes objetivo de precedencia. Se pueden tomar decisiones de marketing más precisas a un grupo. | Se agruparon a los clientes con un comportamiento compartido del consumidor y el valor del cliente; los clientes con más alto valor de consumidor podrían ser la prioridad de clientes objetivo. |
| 12 | Taylor and Francis | Customer behavior and decision making in the refurbishment industry - A data mining approach. | Este artículo puede proporcionar resultados útiles para las empresas innovadoras de renovación en su toma de decisiones. | Los resultados de la investigación muestran que, la mayoría de los encuestados está de acuerdo con que el "servicio de garantía" es una táctica de marketing efectiva y atractiva. | Se muestra que más de la mitad de los clientes tienden a confiar la renovación a empresas de marca. Por lo tanto, hay mucho espacio para el establecimiento de imágenes de marca. |

- | | | | | | |
|----|-----------------------|--|---|--|--|
| 13 | Science Direct | Big data analysis to Features Opinions Extraction of customer. | El objetivo es proponer una arquitectura de big data y extraer opiniones de clientes sobre productos. | Se obtienen las opiniones de los clientes de acuerdo a las características de algunos productos (en este caso, teléfonos celulares). | Se propone y prueba la efectividad del algoritmo con conjuntos de datos disponibles en la web. |
| 14 | IEEE Xplore | Efficient implementation of data mining: Improve customer's behaviour. | El objetivo es ver cómo las técnicas de DM pueden extraer conocimiento de una base de datos grande y cómo analizar el comportamiento de los clientes para mejorar el rendimiento. | En este artículo examinamos la inducción de reglas basadas en datos agrupados. Para extraer reglas o patrones, usamos grupos de clientes basados en la base de datos de clientes en relación con la consulta particular. | El modelo presentado busca mejorar el comportamiento y la comprensión del cliente. La empresa comprenderá los problemas que presentan los clientes y mejorarán su satisfacción. |
| 15 | Science Direct | Application of data mining techniques in customer relationship management: A literature review and classification. | El análisis proporciona una hoja de ruta para guiar investigaciones futuras y facilitar la creación de conocimiento sobre las aplicaciones y técnicas de DM. | Se presentan los diferentes tipos y clasificaciones de los artículos correspondientes a la aplicación de técnicas de DM y sirve para predecir comportamientos futuros, para predecir la efectividad y rentabilidad de la toma de decisiones. | La aplicación de técnicas de minería es una tendencia emergente en la industria. Este artículo brinda información la aplicación y sobre técnicas que se utilizan con mayor frecuencia. |
| 16 | Cuaderno o Activa | Visualizaciones analíticas para la toma de decisiones en pequeñas y medianas empresas utilizando Data Mining. | Realizar visualizaciones analíticas para facilitar el acceso de compañías a información en plataformas de inteligencia de negocios. | Se facilita el acceso a la información a dichas empresas mediante la realización de visualizaciones analíticas. | La información es el recurso más importante de una empresa. De esta manera se toman decisiones relevantes que afecten de manera positiva sus productos y servicios. |
| 17 | Mary Ann Liebert, INC | Data Science and its Relationship to Big Data and Data-Driven Making. | En este artículo se busca demostrar la que es exactamente la ciencia de datos. | | El rendimiento empresarial puede mejorarse sustancialmente a través de la toma de decisiones basadas en datos, tecnologías de big data y técnicas de ciencia de datos. |

18	Google Scholar	Data Mining Techniques: A Source for Consumer Behavior Analysis.	Este estudio busca ayudar a las empresas a mejorar sus estrategias de marketing a través del estudio de sus consumidores.	La minería de asociación encuentra relaciones interesantes entre un gran conjunto de datos, estos pueden ayudar con decisiones comerciales, marketing, etc.	El sistema es útil para el estudio del comportamiento de compra de los clientes y la asociación entre productos que se generan.
19	Research Gate	Mining utility-oriented association rules: An efficient approach based on profit and quantity.	Se propone un enfoque basado en el factor de peso y utilidad para la extracción de reglas de asociación.	Se muestra el rendimiento de los enfoques en la extracción de las reglas de asociación, mostrando resultados efectivos.	Se demostró la efectividad del enfoque propuesto. Puede aplicarse de manera lucrativa para el desarrollo empresarial.
20	IEEE Xplore	Mining food industry's multidimensional data to produce association rules using apriori algorithm as a basis of business strategy.	Se quiere desarrollar una estrategia comercial para abordar el patrón de gusto del cliente.	Se obtienen datos de los productos más altos en reglas de asociación por medio de selección de atributos y fila de selección.	Se deben considerar los vínculos entre atributos, para construir reglas de asociación considerando la presencia de conjuntos de datos.
21	Taylor and Francis	A data mining approach to optimise shelf space allocation in consideration of customer purchase and moving behaviours.	Se propone un método de minería de datos para resolver problemas de asignación de espacio.	Se ilustra un supermercado para demostrar la viabilidad del método propuesto de asignación de espacios.	Se realiza una asociación entre productos principales y menores, buscando reubicar estos últimos para así ponerlos cerca de los principales.
22	Research Gate	Mining of users access behavior for frequent sequential patterns from web logs.	Se explora una nueva técnica de patrón de secuencia frecuente llamada AWAPT.	Esta técnica supera a otros métodos propuestos anteriormente, siendo eficiente y escalable en grandes bases de datos.	Se muestra que el nuevo método extrae el conjunto completo de patrones y es eficiente y más rápido.
23	Google Scholar	Data Science, Predictive Analytics, and Big Data: A Revolution That Will Transform Supply Chain Design and Management.	Se proponen definiciones de ciencia de datos y análisis predictivo aplicados a SCM.	Se proporcionan ejemplos de aplicaciones potenciales de grandes bases de datos dentro de la logística y la práctica de SCM.	Se puede apreciar cómo la ciencia de datos y el análisis predictivo se aplican a SCM.

Fuente de información: Elaboración propia

Al hablar de las características de los artículos según el tipo de documento, podemos observar en Tabla 3 la distribución de los artículos en dos grupos: artículos científicos (17 artículos, 73.91%) que son la mayoría utilizados para este estudio, y artículos de revisión (6 artículos, 26.09%). En cuanto a las características de los artículos según su año de publicación, como se puede observar en Tabla 4, se demostró que entre el 2009 al 2013 se encontraron 3 artículos por cada año (13.04% cada uno), asimismo en el 2017; para los años 2016 y 2018 se redujeron a dos artículos por año (8.70% cada uno), para el 2014 solo se encontró uno (4.35%) y no se encontraron artículos correspondientes al año 2015.

Tabla 3: Características de los estudios según el tipo de documento

Tipo de documento	F	%
Artículo científico	17	73.91
Artículo de revisión	6	26.09
TOTAL	23	100

Fuente de información: Autoría propia

Nota: El tipo de documento de los estudios más encontrado para esta investigación es el artículo científico.

Tabla 4: Características de los estudios según año de publicación

Año de publicación	F	%
2009	3	13.04
2010	3	13.04
2011	3	13.04
2012	3	13.04

2013	3	13.04
2014	1	4.35
2016	2	8.70
2017	3	13.05
2018	2	8.70
TOTAL	23	100

Fuente de información: Autoría propia

Nota: No se registraron artículos para el año 2015.

Con respecto a la tabla 5, se concluye que la mayoría de los artículos encontrados han sido publicados por 22 revistas diferentes (4.55% cada uno), siendo “Procedia Computer Science” la revista que publicó más de un artículo de los utilizados en esta investigación (2 en total, 9.09%); además, se debe tener en cuenta que el artículo “A study of discovering customer value for CRM: Integrating customer lifetime value analysis and data mining techniques” no se incluyó en esta tabla, ya que no cuenta con información de la revista publicadora de dicho artículo.

Tabla 5: Características de los estudios según Revista de publicación

Revista de publicación	F	%
Asia Pacific Journal of Marketing and Logistics	1	4.55
Industrial Management & Data Systems	1	4.55
Journal of Database Marketing & Customer Strategy Management	1	4.55
IJBIR	1	4.55
Procedia Computer Science	2	9.09
International Journal of Research in Marketing	1	4.55
International Journal of Database Theory and Application	1	4.55

Baltic J. Modern Computing	1	4.55
Journal of Consumer Marketing	1	4.55
International Journal of Information and Management Sciences	1	4.55
Journal of Civil Engineering and Management	1	4.55
2009 IEEE/ACS International Conference on Computer Systems and Applications	1	4.55
Expert Systems with Applications	1	4.54
Cuaderno Activa	1	4.54
Foster Provost and Tom Fawcett	1	4.54
Arxiv	1	4.54
International Journal of the Physical Sciences	1	4.54
International Conference of Information and Communication Technology (ICoICT)	1	4.54
International Journal of Production Research	1	4.54
International Journal of Database Management Systems (IIDMS)	1	4.54
Journal of Business Logistics	1	4.54
TOTAL	22	100

Fuente de información: Elaboración propia

Nota: Se registraron solo 22 revistas ya que el artículo de “A study of discovering customer value for CRM: Integrating customer lifetime value analysis and data mining techniques” por Chen, *et al.*, 2009, no tenía información de la revista donde se publicó.

Utilidad de las Herramientas de Data Mining

Uno de los temas más abordados durante esta investigación es el Data Mining (DM) o minería de datos, el cual es útil para estudiar el comportamiento de compra de los clientes en tiendas departamentales minoristas ya que, de esta manera, las compañías descubren la asociación entre los clientes y los diferentes productos de la tienda, es decir, cómo los clientes están cambiando de una marca a otra para satisfacer sus necesidades (Raone y Kullkarni, 2011). Uno de los hallazgos encontrado es que esta información es el recurso más

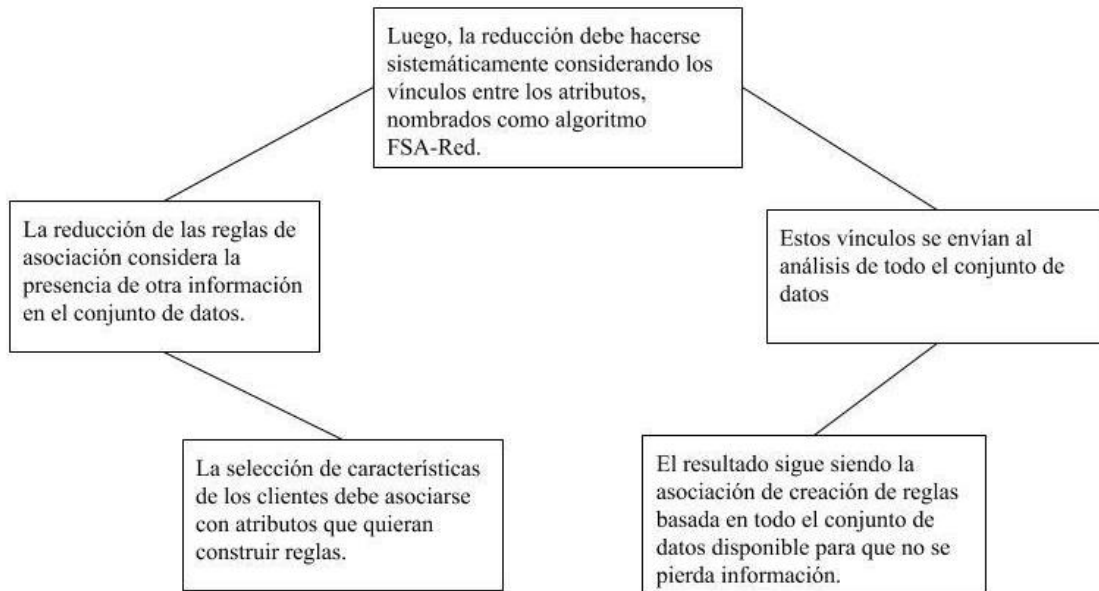
importante para una empresa pues sin él no se podrían ejercer decisiones precisas en la organización (Noreña *et al.*, 2016) ya que existe una fuerte evidencia de que el rendimiento empresarial puede mejorarse a través de la toma de decisiones basadas en información, tecnologías y técnicas de ciencia de datos de Big Data y muchas veces ayudan a automatizar decisiones a escala masiva (Provost y Faucett, 2013). Se encontró que la toma de decisiones siempre ha sido vista como una de las actividades más importantes en una organización, para gerentes hasta empleados, y es difícil sobreestimar el valor de mejorarlo (Davenport, 2010).

Es por eso que, el interés de cuidar este recurso de información, extraerlo y analizarlo permite profundizar en técnicas de DM que permitan su manipulación (Noreña *et al.*, 2016) como lo son las redes neuronales o los árboles de decisión, usados para buscar segmentos rentables de clientes a través del análisis de las características de los clientes (Ngai *et al.*, 2009). El extraer automáticamente estos patrones de datos es un proceso con etapas bien definidas, las cuales se deben comprender porque ayudan a estructurar la resolución de problemas, lo hace más sistemático y, por lo tanto, menos propenso a errores (Provost y Faucett, 2013).

Se halló que hay dos pasos en todo el proceso de DM que son muy cruciales y son los que requieren más tiempo: la preparación de datos e interpretación y evaluación de los modelos (Chen *et al.*, 2012); y que las mejores técnicas para los que no son expertos en DM son las redes neuronales, seguidas por los árboles de decisión y las técnicas de reglas de asociación (Ngai *et al.*, 2009).

Con respecto al caso de reglas de asociación, en la Fig. 1 se describe el proceso de creación de reglas mediante el algoritmo FSA-Red.

Figura 2. Proceso de creación de reglas de asociación usando el algoritmo FSA-Red.



Fuente de información: Sulianta et al., 2013.

Elaboración: Propia.

Luego, se encontró que se recomienda utilizar tres pasos de validación para analizar si las reglas de asociación que se construyeron son confiables para describir todos los datos. Primer paso: entrenamiento de datos después de la reducción; Segundo Paso: entrenamiento de datos sin reducción; Tercer Paso: pruebas de datos (Sulianta *et al.*, 2013). Otra opción hallada para generar un conjunto de reglas de asociación a partir de una base de datos es utilizar el algoritmo tradicional Apriori. Para cada regla de asociación extraída, se calcula una puntuación ponderada de utilidad combinada (UW-Score) en función de las restricciones de peso (W-gain) y utilidad (U-gain) (Sandhu et al., 2011).

Con respecto a los árboles de decisión, se mostraron los resultados experimentales sobre el rendimiento de la herramienta Adaptive Web Access Pattern Tree (AWAPT) en

comparación con WAP Tree y FS-Tree. Muestra que AWAPT supera a otros métodos propuestos ya que es eficiente y escalable para extraer patrones secuenciales en grandes bases de datos, elimina la necesidad de almacenar numerosos árboles WAP intermedios durante la extracción, y reduce drásticamente los enormes costos de acceso a la memoria, especialmente cuando se extraen secuencias muy largas con millones de registros. (Vijayalakshmi *et al.*, 2010).

Al hablar de diseño de organización de artículos en tiendas de retail, se recomienda utilizar el algoritmo High Utility Itemset Mining o Minado de Alta Utilidad (HUIM) para organizar los artículos de manera eficiente (Altuntas, 2017), así como el Location Preference Evaluation (LPE) para calcular la preferencia de ubicación si se coloca un artículo menor en una determinada sección de la tienda (Tsai y Huang, 2014).

Por otro lado, el modelo de Gelbrich & Wünschmann (GVM), que es un modelo de comportamiento del cliente futuro-pasado para segmentar clientes con base en su CLV, arrojó un mejor rendimiento para todas las variables de control, y las segmentaciones de clientes obtenidas a través de este modelo podrían verse más efectivas en comparación con el modelo de Recencia, Frecuencia y Monetario (RFM), que es un modelo de comportamiento del cliente pasado (Hiziroglu *et al.*, 2018).

Por otra parte, se utilizó la técnica Fuzzy Analytical Network Process (FANP), que permite medir la dependencia entre variables RFM y modelar atributos de precisión e imprecisión y se introdujo una metodología novedosa que une FANP y Self Organization Map (SOM) para agrupar a los clientes, se cree que ayuda a la compañía a comprender más fácilmente y enfocar de manera más precisa a los clientes objetivo (Golmah y Mirhashemi, 2012).

Para determinar qué clientes pertenecen al "grupo dorado de clientes objetivo", que son quienes tienen el CLV más alto, y cuáles son sus patrones de comportamiento, se hallaron las recomendaciones de usar las herramientas de DM como las reglas de asociación, On-Line Analytical Processing (OLAP) y árbol de decisión (Chen *et al.*, 2009).

También se dio a conocer que la aplicación de la metodología Cross-Industry Process for Data Mining (CRISP-DM) es estándar ante otras herramientas en la extracción de información en DM, y que en compañía de visualizaciones analíticas, permite observar detalladamente la información, ayudando a que pequeñas y medianas empresas puedan tomar decisiones que los afecten positivamente. Además, se optimizan recursos y se facilitan el análisis de su comportamiento futuro en la empresa (Noreña *et al.*, 2016).

Para el desarrollo de encuestas, otros estudios recomiendan combinar modelos Engel-Kollat-Blackwell (EKB) y factores influyentes relevantes, ya que la investigación del cuestionario demuestra que un pequeño número de preguntas clave son suficientes para medir las actitudes de los clientes, reducir el tiempo de respuesta y aumentar la tasa de recolección (Huang y Hsueh, 2010).

De igual manera, actualmente, hay una proliferación de datos de opinión del consumidor como resultado de Tweets, "Me gusta" y reseñas de productos en sitios web, los cuales deben ser analizados y cuantificados (Waller y Faucett, 2013) extrayendo las características de las opiniones de los clientes mediante tecnologías de Big Data combinadas con herramientas de machine learning y minería de texto, para lo cual se recomienda un algoritmo de determinación de polaridad de opinión basado en MapReduce, para luego poder validar su efectividad y finalmente, usar cámaras de datos de la plataforma de Microblogging Twitter (Mars y Gouider, 2017), sin embargo, se halló que es necesario probar este sistema

con otros puntos de referencia y realizar más investigación en la minería de opinión para la evaluación del comportamiento del consumidor.

Relación entre la ubicación de artículos y el comportamiento de clientes

Se encontró que la ubicación de los artículos en tiendas minoristas afecta directamente las actitudes del consumidor e intenciones de compra, porque ubicar artículos considerando asociaciones entre ellos y los hábitos de consumo de los clientes aumenta la relación de ventas dentro del diseño actual.

Por lo tanto, las empresas pueden organizar los artículos de manera eficiente para aumentar la cantidad de artículos vendidos en negocios minoristas mediante el algoritmo High Utility Itemset Mining (HUIM), basando sus datos en estrategias para asignar familias de productos en el diseño de cada almacén (Altuntas, 2017). Además, en otro estudio se propone un enfoque de DM para resolver los problemas de asignación de espacio en los estantes al considerar relaciones de asociación de productos con clientes y conocimiento de la ruta de compras. De esta manera, le permite a los gerentes de tiendas minoristas optimizar la asignación y el posicionamiento del producto considerando transacciones anteriores, junto con el comportamiento del cliente recopilado mediante tecnología Radio Frequency Identification (RFID) o Identificación de Frecuencia Radial y del procedimiento de Location Preference Evaluation (LPE) para calcular la preferencia de ubicación si se coloca un artículo en una determinada sección de la tienda (Tsai y Huang, 2014).

Se sabe que la empresa puede obtener una mejor comprensión de los consumidores al explorar las asociaciones entre los grupos de consumidores y los productos que han comprado; examinar en qué momentos específicos (temporadas), qué productos y qué tipos

de productos han comprado con frecuencia será beneficioso para el negocio (Chen et al., 2012). Estos datos capturados se deben evaluar con más detalle para un mayor beneficio y así, en lugar de simplemente registrar que una unidad se vendió en un lugar en particular, también se captura la hora en que se vendió y la cantidad de inventario en el momento de la venta (Waller Y Faucett, 2013).

Con respecto a la recomendación de productos, se sabe que es una estrategia madura y aún efectiva en la actualidad, la recopilación de reglas de asociación se puede utilizar para construir dicha estrategia con el fin de determinar la forma en que se mostrará el producto o si se formarían paquetes para alcanzar la meta de ventas (Sulianta y Atastina, 2013).

Aplicación de Data Mining en Customer Lifetime Value

El uso de modelos Customer Lifetime Value (CLV) o Valor del Tiempo de Vida del Cliente y técnicas de DM en conjunto brinda una enorme capacidad a las organizaciones para reconocer grupos de clientes de alto valor, de esta manera, el CLV y los conceptos de segmentación dentro del contexto de data mining pueden proporcionar una comprensión práctica en el área de análisis del cliente (Hiziroglu *et al.*, 2018).

Se propone combinar estos conceptos ya que permiten analizar aún más los datos de transacciones y datos sociodemográficos para empíricamente descubrir el patrón de compra de clientes, sus características donde hay información que es relevante en todas las etapas del ciclo de toma de decisiones del consumidor como lo que hace, cómo se hace, dónde consume, cuándo lo hace y con quién consume; y segmentos con base en el Valor Actual de los Clientes (CCV) y Valor Potencial de los Clientes (CPV) (Chen *et al.*, 2009; Hofacker *et al.*, 2016). De esta manera se pueden clasificar sus valores basándose en función del cambio estimado en su CLV mediante dos métricas: la correlación intra-clientes, que proporciona

información sobre la volatilidad de los flujos de efectivo generados por un cliente, y la correlación inter-clientes, que proporciona información en la compensación entre la magnitud del flujo de ingresos generado por los clientes (tiempo x gasto) y la durabilidad esperada de este flujo (abandono) (Glady *et al.*, 2014).

Así, el concepto de CLV ayuda a los que toman decisiones a enfocarse en los mercados con mayor claridad en entornos fuertemente competitivos (Golma y Mirhashemi, 2012); debido a que se conoce que monitorear la diversidad del grupo de clientes más diverso y predecir qué cliente se afiliará potencialmente al grupo más rentable o al menos rentable es muy útil para el negocio a largo plazo (Chen *et al.*, 2012).

Aplicación de Data Mining en Customer Relationship Management

Más allá de simplemente comprender el valor del cliente, la compañía intenta establecer una mejor relación con el cliente, aumentar la lealtad y los ingresos del cliente (Hsieh y Chu, 2009). Para ello, Customer Relationship Management (CRM) o Gestión de Relaciones con Clientes es una estrategia recomendada para aprender más sobre las necesidades y comportamientos de los clientes para desarrollar relaciones más sólidas con ellos.

El modelo más aplicado en CRM es el modelo de clasificación, el cual sirve para predecir comportamientos futuros del cliente ya que puede identificar el comportamiento del cliente y aumentar su lealtad hacia una organización (Ngai *et al.*, 2009). Igualmente, la técnica de inducción de reglas y la técnica de agrupación en los datos del cliente es un factor clave para comprender mejor a los clientes y para la mejora de CRM en cualquier organización (Mudimigh *et al.*, 2009).

Por otro lado, la gestión de reclamaciones es un requisito crucial para las empresas exitosas cuando gestiona las necesidades de los clientes y los cambios en su comportamiento. Las técnicas de DM podrían aplicarse para descubrir patrones invisibles de quejas de la Big Data de una empresa con el fin de mejorar la relación con sus clientes (Ngai *et al.*, 2009).

Finalmente, después del análisis completo y la evaluación, se puede incluir la acción de la organización para obtener una respuesta positiva con respecto a la consulta particular del cliente, la predicción de la escalada de ventas, un nuevo plan de marketing, nuevas estrategias de publicidad e instrucciones para sus empleados (Mudimigh *et al.*, 2009); para concluir, se puede decir que la investigación sobre la aplicación de DM en CRM aumentará significativamente en el futuro en función de las tasas de publicación anteriores y el creciente interés en el área. (Ngai *et al.*, 2009).

Importancia de Data Mining en Estrategias de Marketing

Muchas veces las empresas de la industria hacen más hincapié en la tecnología y el mantenimiento de los negocios, mas no en el concepto de "marketing" y cómo satisfacer a los consumidores, es por esto, que se necesita explorar las características del comportamiento del consumidor con teorías y métodos relevantes para proporcionar un acceso a las percepciones, pensamientos y requisitos de los clientes de modo que las empresas puedan ajustar sus políticas de marketing y modelos de comercialización para cumplir con las expectativas de los clientes (Huang y Hsueh, 2010).

Con este fin, se agrupan a los clientes que compartan comportamiento y CLV para después revisar sus perfiles utilizando el árbol de decisiones. Se demostró que los clientes con más alto CLV podrían ser la prioridad de clientes objetivo (Hsieh y Chu, 2009); y a su

vez, pueden ayudar a la empresa a comprenderlos mejor en términos de su rentabilidad y, en consecuencia, adoptar políticas de marketing apropiadas para los diferentes consumidores (Chen *et al.*, 2012).

En adición, los gerentes deben ser conscientes de las preocupaciones de privacidad del consumidor en la recopilación y utilización de Big Data para estos fines, incluso cuando lo usan para entregar productos y promociones más relevantes para ellos (Hofacker *et al.*, 2016).

Comportamiento de consumidores en tiendas minoristas online

Finalmente, al hablar del comportamiento de clientes en tiendas de retail online se observó que es un tema poco investigado, pero con mucho potencial para profundizar en su investigación debido al aumento del comercio electrónico. En consecuencia, se reveló que existe una correlación entre el alto número de visitas a ciertos productos y sus ventas, ya que a medida que aumentaba la cantidad de búsquedas de ciertos productos por parte del cliente, también aumentaba la posibilidad de que se compraran ciertos productos, además, también se reveló que a medida que aumenta la duración de las compras por parte de los clientes, la cantidad total de productos que se compran al mismo tiempo también aumenta.

A medida que la interacción y la calidad de las compras a través de la herramienta Digital Signage-Based Online Store (DSOS) aumentara, la relación de los clientes con el minorista se fortaleció, lo que ayudó a elevar los ingresos del negocio logrando el objetivo clave de CRM.

Se demostró que el sistema propuesto era capaz de procesar una gran cantidad de datos de entrada fácilmente con un aumento en el número de clientes en el DSOS. (Alfian *et*

al., 2018). De esta manera, se propuso un proceso todo en uno para recopilar y estructurar datos del contenido, la estructura y los usuarios de un sitio web de comercio electrónico. Este proceso se enriqueció con el cruce de los datos recopilados de diversas fuentes para encontrar relaciones no explícitas (Dias y Ferreira, 2017).

CAPÍTULO IV. DISCUSIONES Y CONCLUSIONES

La revisión de la literatura científica realizada permitió encontrar un total de 23 artículos relacionados a data mining y comportamiento del consumidor, publicados entre el 2009 y 2019. Esto permitió resumir en una sola investigación cómo se relaciona el Data Mining para conocer el comportamiento del consumidor y qué técnicas son las comúnmente usadas. Esta investigación no pretende ser exhaustiva, pero proporciona información importante e incidente sobre este tema.

Considerando lo expuesto, se puede afirmar que para entender el comportamiento del consumidor se debe minar: i) Sus patrones de compra como la ruta que toma cuando realiza compras, temporadas en que compra ciertos productos, cantidades de cada producto comprado, etc. Además de sus preferencias, con el fin de organizar los productos en la tienda de una manera más eficiente basándose en el comportamiento de los clientes y realizar promociones acorde al perfil de los mismos; ii) El Customer Lifetime Value, ya que esta valoración permite comprender el valor del cliente para la empresa y separarlos en segmentos de mayor a menor aporte para la compañía y de esta manera, enfocarse en los mercados de mayor rentabilidad con mayor claridad; iii) Sus necesidades, quejas y cambios en su comportamiento mediante la estrategia de Customer Relationship Management para mejorar su relación con ellos, aumentar su lealtad e ingresos hacia la organización; iv) Sus percepciones, pensamientos y requisitos para aplicar estrategias de marketing más precisas a los clientes objetivos de mayor prioridad; v) Qué productos buscan y comparan en tiendas de retail online para aumentar la cantidad de productos comprados por cliente, tema del cual se tiene poca información, por lo que se recomienda ahondar a detalle para otras investigaciones en caso sea necesario. También se puede afirmar que existen diferentes

técnicas de DM para hacer estos análisis del comportamiento, dependiendo del tema en que se quiera minar con profundidad, sin embargo, se recomiendan las redes neuronales, seguidas por los árboles de decisión y las técnicas de reglas de asociación para los que no son expertos en DM. (Chen *et al.*, 2012; Waller Y Faucett, 2013; Golma y Mirhashemi, 2012; Ngai *et al.*, 2009; Huang y Hsueh, 2010; Alfian *et al.*, 2018).

Este estudio es importante porque es un tema que aún está implementando en las empresas y muchas personas aún no conocen o difícil de comprender para algunos. No obstante, este estudio presenta algunas limitaciones como que esta investigación solo se limitó a nueve bases de información, que son las más conocidas e importantes, sin embargo, debe haber muchas más que tengan artículos que no se han revisado. Por otro lado, solo se encontró un artículo en español, la gran mayoría eran artículos en inglés por lo que se cree importante que más investigaciones científicas sobre Data Mining y comportamiento del consumidor deben discutidas y publicadas en otros idiomas. Por ello, se recomienda para próximas investigaciones, realizar la búsqueda de información en más de un idioma y no limitarse solo al español.

REFERENCIAS

Alfian, G., Ijaz, M. F., Syafrudin, M., Syaekhoni, M. A., Fitriyani, N. L. y Rhee, J. (2018).

Customer behavior analysis using real-time data processing: A case study of digital signage-based online stores. *Asia Pacific Journal of Marketing and Logistics*, 31(1), 265-290. <https://doi.org/10.1108/APJML-03-2018-0088>

Altuntas, S. (2017). A novel approach based on utility mining for store layout: a case

study in a supermarket. *Industrial Management & Data Systems*, 117(2), 304-319. <https://doi.org/10.1108/IMDS-01-2016-0040>

Chen, C. W., Yang, C. y Lin, C. S. (2009). A study of discovering customer value for CRM: Integrating customer lifetime value analysis and data mining techniques.

<https://doi.org/0.30136/ZXYGLKX.200907.0002>

Chen, D. Sain, S. L. y Guo, K. (2012). Data mining for the online retail industry: A case study of RFM model-based customer segmentation using data mining. *Journal of Database Marketing & Customer Strategy Management*, 19(1), 197–208.

<https://doi.org/10.1057/dbm.2012.17>

Davenport, T. H. (2010). Business Intelligence and Organizational Decisions. *IJBIR*,

1(1), 1-12. <https://doi.org/10.4018/jbir.2010071701>

Dias, J. y Ferreira, H. (2017). Automating the Extraction of Static Content and Dynamic Behaviour from e-Commerce Websites. *Procedia Computer Science*, 109(1), 297-304. <https://doi.org/10.1016/j.procs.2017.05.355>

Glady, N., Lemmens, A. y Croux, C. (2014). Unveiling the relationship between the transaction timing, spending and dropout behavior of customers. *International Journal of Research in Marketing*, 32(1), 78-93. <https://doi.org/10.1016/j.ijresmar.2014.09.005>

Golmah, V. y Mirhashemi, G. (2012). Implementing A Data Mining Solution To Customer Segmentation For Decayable Products - A Case Study For A Textile Firm. *International Journal of Database Theory and Application*, 5(3), 73-90. Recuperado de: <https://www.earticle.net/Article/A207852>

Hiziroglu, A., Sisci, M., Cebeci, H. I. y Seymen, O. F. (2018). An Empirical Assessment of Customer Lifetime Value Models within Data Mining. *Baltic J. Modern Computing*, 6(4), 434-448. <https://doi.org/10.22364/bjmc.2018.6.4.08>

Hofacker, C. F., Malthouse, E. C. y Sultan. (2016). Big Data and consumer behavior: imminent opportunities. *Journal of Consumer Marketing*, 33(2), 89-97. <https://doi.org/10.1108/JCM-04-2015-1399>

- Hsieh, N. C., y Chu, K. C. (2009). Enhancing Consumer Behavior Analysis by Data Mining Techniques. *International Journal of Information and Management Sciences*, 20(1), 39-53. Recuperado de: jims.ms.tku.edu.tw/list.asp
- Huang, C. F. y Hsueh, S. L. (2010). Customer behavior and decision making in the refurbishment industry - A data mining approach. *Journal of Civil Engineering and Management*, 16(1), 75-84. <https://doi.org/10.3846/jcem.2010.07>
- Mars, A. y Gouider, M. (2017). Big data analysis to Features Opinions Extraction of customer. *Procedia Computer Science*, 112(1), 906-916. <https://doi.org/10.1016/j.procs.2017.08.114>
- Mudimigh, A. A., Saleem, F. y Ullah, Z. (2009). Efficient implementation of data mining: Improve customer's behaviour. *2009 IEEE/ACS International Conference on Computer Systems and Applications*, 1(1), 7-10. <https://doi.org/10.1109/AICCSA.2009.5069289>
- Ngai, E.W.T., Xiu, L. y Chau, D.C.K. (2009). Application of data mining techniques in customer relationship management: A literature review and classification. *Expert Systems with Applications*, 36(2), 2592-2602. <https://doi.org/10.1016/j.eswa.2008.02.021>
- Noreña, P., Quiceno, G. y Uribe, L. (2016). Visualizaciones analíticas para la toma de decisiones en pequeñas y medianas empresas utilizando Data Mining. *Cuaderno*

Activa, 8(1), 31-40. Recuperado de

<https://ojs.tdea.edu.co/index.php/cuadernoactiva/article/view/328>

Provost, F. y Fawcett, T. (2013). Data Science and its Relationship to Big Data and Data-Driven Decision Making. *Foster Provost and Tom Fawcett*, 1(1), 51-59.

<http://doi.org/10.1089/big.2013.1508>

Raorane, A. y Kulkarni, R. V. (2011). Data Mining Techniques: A Source for Consumer Behavior Analysis. *Arxiv*, 1(1). <https://doi.org/0.5121/ijdms.2011.3304>

Sandhu, P. S., Dhaliwal, D. S., Panda, S. N. (2011). Mining utility-oriented association rules: An efficient approach based on profit and quantity. *International Journal of the Physical Sciences*, 6(2), 301-307. <https://doi.org/10.5897/IJPS09.303>

Sulianta, F., Liong, T. H., & Atastina, I. (2013). Mining food industry's multidimensional data to produce association rules using apriori algorithm as a basis of business strategy. *2013 International Conference of Information and Communication Technology (ICoICT)*, 1(1), 176-181,

<https://doi.org/10.1109/ICoICT.2013.6574569>

Tsai, C.-Y., Huang, S.-H. (2014). A data mining approach to optimise shelf space allocation in consideration of customer purchase and moving behaviours.

International Journal of Production Research, 53(3), 850-866.

<https://doi.org/10.1080/00207543.2014.937011>

Vijayalakshmi, S., Mohan, V., Suresh Raja, S. (2010). Mining of users access behavior for frequent sequential pattern from web logs. *International Journal of Database Management Systems (IJDMS)*, 2(3), 31-45.
<https://doi.org/10.5121/ijdms.2010.2304>

Waller, M., Fawcett, S. (2013). Data Science, Predictive Analytics, and Big Data: A Revolution That Will Transform Supply Chain Design and Management. *Journal of Business Logistics*, 34(2), 77-84. <https://doi.org/10.1111/jbl.12010>