

Aplicación de la ciencia de datos en la toma de decisiones empresariales en el sector retail colombiano

Application of Data Science in Business Decision-Making in the Colombian Retail Sector

Victor Andrés Lasso Vivas¹

Universidad Nacional Abierta y a Distancia, Colombia

Resumen.

En un entorno altamente competitivo, cambiante y saturado, los minoristas del país se enfrentan al desafío de obtener ventajas y tomar medidas para evitar pérdidas o incluso luchar contra la desaparición. Esta investigación tiene como objetivo demostrar cómo las capacidades de la ciencia de datos y las nuevas herramientas de análisis de datos contribuyen a la toma de decisiones informadas, mejora de la eficiencia y rentabilidad de los negocios minoristas a partir del aprovechamiento de los datos generados en sus operaciones diarias. Esta investigación se construye a partir de una revisión sistemática de la literatura donde se examinan los elementos metodológicos, técnicos y tecnológicos que definen la ciencia de datos. Se discuten las ventajas y posibilidades que la aplicación de la ciencia de datos puede ofrecer al sector retail en términos de toma de decisiones estratégicas y se considerarán posibles proyecciones y se evaluará el impacto potencial de la ciencia de datos en el futuro del sector retail. Los resultados de esta investigación brindan información valiosa sobre el sector minorista y las dificultades que enfrenta en la toma de decisiones. Además, se exploran las tendencias actuales en la aplicación de la ciencia de datos para la toma de decisiones en distintas áreas incluyendo el sector minorista, analizando las tácticas, tecnologías y resultados de estas aplicaciones. Finalmente, se logra identificar y destacar las ventajas, oportunidades y beneficios que surgen de la aplicación de la ciencia de datos en la toma de decisiones en el sector minorista.

Palabras clave:

ciencia de datos, toma de decisiones, big data, comercio, minorista, inteligencia artificial

Abstract

In a highly competitive, changing, and saturated environment, retailers in the country face the challenge of gaining advantages and taking measures to prevent losses or even fight

¹ Maestría en Gestión de tecnologías de la información, <https://orcid.org/0000-0002-1566-3072/>
valassov@unadvirtual.edu.co

against extinction. This research aims to demonstrate how data science capabilities and new data analysis tools contribute to informed decision-making, improved efficiency, and profitability of retail businesses by leveraging the data generated in their daily operations. This research is built upon a systematic literature review that examines the methodological, technical, and technological elements that define data science. The advantages and possibilities that the application of data science can offer to the retail sector in terms of strategic decision-making are discussed, and possible projections are considered, evaluating the potential impact of data science on the future of the retail sector. The results of this research provide valuable information about the retail sector and the difficulties it faces in decision-making. Furthermore, current trends in the application of data science for decision-making in various areas, including the retail sector, are explored, analyzing the tactics, technologies, and outcomes of these applications. Finally, the advantages, opportunities, and benefits that arise from the application of data science in decision-making in the retail sector are identified and highlighted.

Keywords: Data Science, Decision Making, Big Data, Commerce, Retail, Artificial Intelligence.

1. Introducción

El sector minorista desempeña un papel fundamental en la economía del país reflejada en su participación en el PIB: 12 %, superado solamente por la industria 15 % y sector financiero 14 % (Navas Ríos *et al.*, 2004). Esta situación se atribuye, en parte, a la arraigada cultura de los colombianos de realizar sus compras en las tiendas de barrio, además, al bajo nivel de ingresos y la dispersión de la población urbana (Navas Ríos *et al.*, 2004). Como resultado, el sector minorista se ha consolidado como una opción conveniente y accesible para los consumidores en el país.

En el contexto empresarial los términos “minorista” y “retail” se utilizan comúnmente para referirse a la misma actividad comercial. Según la definición técnica del comercio minorista, se refiere a la reventa (compra y venta sin transformación) de mercancías o productos, destinados para consumo o uso personal o doméstico (consumidor final) (DANE, 2019). Bajo este concepto, el mercado minorista agrupa comerciantes, distribuidores o prestadores de servicios que concentran el desarrollo de sus actividades en el consumidor final (Martínez-Daza, 2022).

El entorno dentro del que se mueve el comercio minorista involucra las altas tasas de desempleo (alrededor del 11 %) (DANE, 2023) conducentes al crecimiento del comercio informal y las altas cargas tributarias (Colombia paga una de las tasas de tributación más altas de Latinoamérica con una tasa de crecimiento alrededor del 2,5 %) (Navas Ríos *et al.*, 2004). Sin embargo, ante estas dificultades del entorno existen oportunidades de aumentar el fortalecimiento en la administración del negocio como posibilidades de acceso a crédito; plazos especiales para el pago de impuestos; y amnistías en los reportes ante centrales de

riesgo (Fenalco, 2021a). Los retos a los que en verdad se enfrentan los comercios minoristas se sintetizan en una alta competencia y problemas con el suministro de productos (Fenalco, 2021b). De esta manera los dueños y gerentes se enfrentan a retos que deben resolver con la toma de decisiones con base en datos que ofrezcan la certeza y confianza a los comercios minoristas de que esas decisiones ofrecerán un margen mínimo de error. Como menciona Molina Azorín *et al.* (2012), el objetivo de esta toma de decisiones es encontrar soluciones óptimas en cada situación, considerando las características propias de la organización.

En este contexto, surge el interés por aplicar disciplinas como la ciencia de datos. La ciencia de datos es un área interdisciplinaria que tiene como propósito transformar datos en valor para poder reportar, diagnosticar, predecir y también recomendar soluciones o mejoras en productos, servicios y/o procesos (Arriagada-Benítez, 2020). Los autores que hablan del ciclo de vida en proyectos de ciencia de datos lo suelen dividir en 6 fases: descubrimiento, preparación de datos, planeación del modelo, construcción del modelo, operación y comunicación de resultados. En general, para la mayoría de los autores estas 6 fases componen el ciclo de vida, aunque otros señalan que se puede describir en 5 fases (Jain & Kushagra, 2022) o, en los marcos ajustados a la extracción de información para aplicaciones empresariales, los autores usan 4 etapas (Singh Yadav, 2022). Al revisar y comparar las fases de estos ciclos de vida se puede ver que terminan en el uso de los datos, pero no se enfoca en un uso específico de resultados como la toma de decisiones. Aquí es donde entran metodologías como CRISP-DM que, en una de sus etapas, la etapa evaluación, incluye la evaluación del modelo y su impacto en la toma de decisiones. Si bien existen otras metodologías usadas para guiar el proceso de análisis de datos y el desarrollo de modelos en el campo de la ciencia de datos, como KDD (Knowledge Discovery in Databases) y SEMMA (Sample, Explore, Modify, Model, Assess) no proporcionan un marco tan completo como CRIPS-DM, la cual está principalmente orientada a la empresa y se centra específicamente en el proceso de minería de datos y descubrimiento de conocimiento a partir de los datos (Wirth & Hipp, 2000).

Apoyada en esta metodología CRIPS-DM, la Unidad de Científicos de Datos (UCD) de Colombia, propone un modelo de analítica de datos para la para la toma de decisiones en el sector público (Botero Barco, *et al.*, 2021).

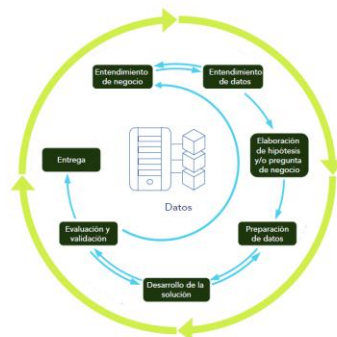


Figura 1. Etapas para el desarrollo de un proyecto de analítica de datos. Adaptado de metodología CRISP-DM (Botero Barco, et al., 2021).

En este modelo se propone iniciar en el entendimiento del negocio y elaborar una hipótesis o pregunta de negocio a la que se va a dar respuesta con el desarrollo del proyecto. De esta manera resulta un marco en el que se tienen en cuenta etapas importantes de la metodología CRISP-DM como la comprensión del negocio y la elaboración de hipótesis o pregunta de negocio, y se tienen en cuenta etapas importantes del ciclo de vida de ciencia de datos como el entendimiento de los datos, la preparación de los datos, la creación del modelo y la evaluación de los resultados.

2. Estrategias y técnicas de aplicación ciencia de datos

La ciencia de datos permite ser aplicada en infinidad de situaciones en las que se necesite tomar decisiones, por ejemplo, en campos de la ciencia y la ingeniería (Montáns *et al.*, 2019), en entornos militares (Heilman *et al.*, 2019), en la salud (Wu & Xiao, 2021), decisiones comerciales (Coussement & Benoit, 2021), entre otras. Las aplicaciones de ciencia de datos para la toma de decisiones varían desde el uso en análisis de información, realizar predicciones (Schnepf *et al.*, 2022) y recomendar acciones a partir de datos. En (Rajesh *et al.*, 2020) usa la ciencia de datos para la toma de decisiones relacionadas a la aprobación de préstamos bancarios desde la perspectiva de un abogado. Esta toma de decisión se hace por medio de modelo de árbol de decisiones en el que se ajustan pesos a las características más importantes que se consideran al efectuar préstamos bancarios, estos pesos se fijan por medio de un sistema de predicción automatizado. En Veres *et al.* (2021) se desarrolló un modelo para la toma de decisiones sobre la financiación de proyectos. Para hacerlo el autor propone un método para la toma de decisiones que consta de varios pasos, incluyendo la recolección de datos relevantes, el análisis exploratorio de datos, la selección y preparación de los datos, la selección del modelo adecuado, el entrenamiento y ajuste del modelo, y la evaluación y validación del modelo. Una vez que se ha desarrollado el modelo, se puede utilizar para predecir los resultados financieros potenciales de nuevos proyectos de financiamiento, lo que permite a los inversores tomar decisiones más informadas y reducir el riesgo de pérdidas financieras.

La técnica de ciencia de datos más utilizada para la toma de decisiones en estas investigaciones es la minería de datos (Schnepf *et al.*, 2022; Rajesh *et al.*, 2020), sobre todo debido a que es una técnica no supervisada lo que significa que no requiere una hipótesis previa sobre el valor de los datos y así se pueden descubrir patrones y relaciones. Aunque también se suelen usar técnicas de aprendizaje automático y herramientas de análisis estadístico como la optimización bayesiana secuencial, CB-SEM y regresión (Van der Voort *et al.*, 2021). Otro aspecto para destacar es que la herramienta más usada en estas aplicaciones suele ser Python (permite el uso de múltiples librerías) y en cuanto al modelo de análisis suele usarse modelos de árbol de decisión (Rajesh *et al.*, 2020; Van der Voort *et al.*, 2021) o modelos de clasificación (Sarlis *et al.*, 2021).

Gracias a las técnicas y modelos utilizados en ciencia de datos, se abren diversas posibilidades estratégicas para la toma de decisiones dinámicas como el cambio de un precio, variedad de productos que se tiene en venta, publicidad o decisiones de alto riesgo como la apertura de nuevas tiendas, ampliaciones, renovaciones de tiendas y adquisiciones. Para las decisiones dinámicas se tiene aplicabilidad en la estimación de costes (Fernández-Revuelta Pérez & Romero Blasco, 2022), el análisis del manejo del surtido, la planificación de la disposición de los productos en los estantes, análisis de fidelidad, estimación medios de pagos de los clientes, el pronóstico de ventas (Toro Ocampo *et al.*, 2004; Ma & Fildes, 2021), la gestión del inventario y la segmentación de clientes (Cam Gensollen, 2022; Silva Guerra, 2012). Por otro lado, en respuesta a las decisiones de alto riesgo como la expansión física de las empresas de comercio minorista, resulta importante considerar datos para elegir la localización óptima de los nuevos puntos de venta ya que este es un asunto que podría condicionar la sostenibilidad en el tiempo de estas empresas. La ciencia de datos a través de la utilización de big data se emplea para estimar la compatibilidad demográfica entre las empresas de comercio minorista y los centros comerciales, que además permite medir el desempeño agregado de los negocios minoristas y de sus diferentes establecimientos (Gutiérrez Barrera, 2018).

3. Discusión caso de aplicación sector retail

Al analizar los datos históricos y utilizar técnicas de modelado predictivo, se pueden simular diferentes escenarios y pronosticar los resultados. Esto reduce la incertidumbre y permite tomar decisiones más informadas y fundamentadas. De esta manera los minoristas pueden acercarse a los resultados esperados al ajustar su estrategia y tácticas para maximizar las oportunidades de éxito, esto les permite adaptarse rápidamente a los cambios en el mercado, las preferencias del cliente y las condiciones económicas. Por otro lado, la clasificación detallada de los clientes en distintos grupos proporciona beneficios significativos para la toma de decisiones en un minimercado. Al analizar cada grupo clasificado, se pueden obtener resultados reveladores que arrojan luces sobre las preferencias y comportamientos de los clientes. De igual manera este conocimiento del cliente proporciona a los minimercados una ventaja competitiva significativa al permitir identificar en qué productos los consumidores priorizan exclusivamente el factor precio, dejando de lado consideraciones como el prestigio de la marca, la calidad u otras características adicionales. Esto les brinda una oportunidad única para adaptar su estrategia y ofrecer productos que se ajusten perfectamente a las preferencias y necesidades de sus clientes, asegurando así su satisfacción y fidelidad. Emplear la ciencia de datos para la toma de decisiones implica un cambio de paradigma en todo el negocio. Promueve una estrategia fundamentada en datos, ya que la ciencia de datos puede ser utilizada para respaldar un análisis de FODA, al identificar riesgos y oportunidades, identificar patrones no detectados con anterioridad, así como evaluar tendencias pasadas y futuras para llegar a una estrategia sólida y efectiva lo que impulsará a los negocios a pensar fuera de la caja, desarrollar una cultura de innovación y experimentación tras entender el alcance de estas herramientas disruptivas. Así mismo, al ofrecer la posibilidad de empoderar a los líderes empresariales para tomar decisiones

fundamentadas en evidencia cuantificable y basada en datos, la ciencia de datos conduce hacia un modelo de autogestión de la información y mejora continua de las estrategias internas de las organizaciones (Rodríguez, 2016).

4. Conclusiones

La falta de conocimiento sobre los datos y la forma de utilizarlos adecuadamente es un problema común en los negocios minoristas. Muchos gerentes y encargados aún dependen de la intuición y la experiencia personal en lugar de utilizar enfoques basados en datos. El empleo de técnicas y herramientas de ciencia de datos como el análisis predictivo, aprendizaje automático y minería de datos como apoyo a la toma de decisiones, permite aprovechar los datos generados en las operaciones diarias de los negocios minoristas para comprender y aprovechar las tendencias actuales en el mercado, identificar oportunidades de mejora y optimización en áreas como la gestión de inventario, la segmentación de clientes y la personalización de la experiencia de compra. Al utilizar enfoques basados en datos, se puede obtener una comprensión más profunda y precisa de los patrones y comportamientos del mercado, lo que facilita la toma de decisiones informadas.

Es importante que el proceso de uso de la información para la toma de decisiones esté alineado con los objetivos comerciales y se obtengan resultados relevantes. En este sentido, la metodología CRISP-DM (Cross Industry Standard Process for Data Mining) resulta útil, ya que está principalmente orientada a la empresa y se centra específicamente en el proceso de minería de datos y descubrimiento de conocimiento a partir de los datos. CRISP-DM abarca etapas importantes como la comprensión del negocio, la preparación de los datos, la creación del modelo y la evaluación de los resultados, lo que garantiza un enfoque completo y sistemático. La aplicación de la ciencia de datos en el sector minorista tiene el potencial de generar beneficios significativos, como la mejora de la eficiencia operativa, el aumento de la rentabilidad y la optimización de la toma de decisiones estratégicas. Al aprovechar los datos disponibles y utilizar enfoques basados en evidencia, los negocios minoristas pueden tomar decisiones más fundamentadas y obtener una ventaja competitiva en un entorno cada vez más complejo y dinámico.

Referencias

- Arriagada-Benítez, M. (2020). Ciencia de datos: hacia la automatización de las decisiones. *Ingeniare. Revista Chilena de Ingeniería*, 28(4), 556-557, <https://doi.org/10.4067/S0718-33052020000400556>
- Botero Barco, A., Rodríguez Ospino, L. A., Gómez Gaviria, D., García Montaña, A., Ríos García, D. P., Durán Pabón, I. M. & Vanegas Barrero, V. (2021). *Aprovechamiento de datos para la toma de decisiones en el sector público*. Departamento Nacional de Planeación, Bogotá, D.C.

- Cam Gensollen, C. R. (2022). Big data en el mundo del retail: segmentación de clientes y sistema de recomendación en una cadena de supermercados de Europa. *Ingeniería Industrial, edición especial*, 189-216. https://revistas.ulima.edu.pe/index.php/Ingenieria_industrial/article/view/5808
- Coussement, K. & Benoit, F. (2021). Interpretable data science for decision making. *Decision Support Systems*, 150, 113664. <https://doi.org/10.1016/j.dss.2021.113664>
- DANE (2019). Muestra mensual de comercio al por menor –MMCM–. Departamento Administrativo Nacional de Estadística. https://www.dane.gov.co/files/investigaciones/fichas/comercio_servicios/ficha_mmcm.pdf
- DANE (2023). Boletín Técnico Gran Encuesta Integrada de Hogares (GEIH), Principales indicadores del mercado laboral. https://www.dane.gov.co/files/investigaciones/boletines/ech/ech/bol_empleo_mar_23.pdf
- Fenalco (2021a). La tienda de barrio aliado estratégico de la industria y el consumidor final. Perspectiva de Fenalco frente a la reactivación económica. *Fenalco*. <https://acortar.link/DYZ0mi>
- Fenalco (2021b). La tienda de barrio sigue siendo la joya de la corona para los productos de consumo masivo. *Fenalco*. <https://erp.fenalco.com.co/blog/gremial-4/la-tienda-de-barrio-sigue-siendo-la-joya-de-la-corona-para-los-productos-de-consumo-masivo-456>
- Fernández-Revuelta Pérez, L. & Romero Blasco, A. (2022). Un enfoque de ciencia de datos para la toma de decisiones en la estimación de costes - Big Data y aprendizaje automático. *Revista de Contabilidad*, 25(1), 45–57.
- Gutiérrez Barrera, C. A. (2018). *Planes de expansión del comercio minorista: un acercamiento desde la analítica, grandes conjuntos de datos e información de transacciones financieras*. Universidad Eafit. <http://hdl.handle.net/10784/12887>
- Heilman E. *et al.* (2019). Application of Data Science within the Army Intelligence Warfighting Function: Problem Summary and Key Findings. *2019 International Conference on Computational Science and Computational Intelligence (CSCI)*, Las Vegas, NV, USA, (pp. 191-195). <https://doi.org/10.1109/CSCI49370.2019.00039>
- Jain, S. & Kushagra (2022). Comprehensive Survey on Data science, Lifecycle, Tools and its Research Issues. *2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON)*, Faridabad, India (pp. 838-842). <https://doi.org/10.1109/COM-IT-CON54601.2022.9850751>
- Ma, S. & Fildes, R. (2021). Retail sales forecasting with meta-learning. *European Journal of Operational Research*, 288(1), 111-128. <https://doi.org/10.1016/j.ejor.2020.05.038>

- Martínez-Daza, M. A. (2022). The management towards organizational change: The situation of retail shops. *revVISUAL*, 12(4), 1–14.
- Molina Azorín, J. F., López Gamero, M. D., Pereira Moliner, J., Pertusa Ortega, E. M. & Tarí Guilló, J. J. (2012). Métodos híbridos de investigación y dirección de empresas: ventajas e implicaciones. *Cuadernos de Economía y Dirección de la Empresa*, 15(2). 55-62.
- Montáns, F. J., Chinesta, F., Gómez-Bombarelli, R. & Nathan Kutz, J. (2019). Data-driven modeling and learning in science and engineering. *Comptes Rendus Mécanique*, 347(11), 845-855. <https://doi.org/10.1016/j.crme.2019.11.009>
- Navas. M. E., Aldana, E. L. & Madrid J. R. (2004). Evolución del gran comercio minorista en Colombia y sus prácticas contemporáneas adaptativas. *Folletos Gerenciales*, 8(11).
- Rajesh, D. P., Alam, M., Tahernezehadi, M., Vikram, C. & Phaneendra, P. N. (2020). Real Time Data Science Decision Tree Approach to Approve Bank Loan from Lawyer's Perspective. *2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA)*, Miami, FL, USA, (pp. 921-929). <https://doi.org/10.1109/ICMLA51294.2020.00150>
- Rodríguez, J. F. (2016). *Implementación de big data en las organizaciones como estrategia de aprovechamiento de la información para incorporarla a la cadena de valor del negocio*. (Trabajo de grado). Universidad Militar Nueva Granada. <http://hdl.handle.net/10654/14411>
- Sarlis. V., Chatziilias, V., Tjortjis, C. & Mandalidis, D. (2021). A Data Science approach analysing the Impact of Injuries on Basketball Player and Team Performance. *Information Systems*, 99, 101750. <https://www.sciencedirect.com/science/article/abs/pii/S030643792100020X>
- Schnepf, J., Vetter, P., Temel, T., Scheuermann, B. & Schmidt-Thieme, L. (2022). On the Potential of Using ERP Business and System Data for Fraud Detection. *2022 IEEE International Conference on Big Data (Big Data)*, Osaka, Japan, (pp. 3081-3091). <https://doi.org/10.1109/BigData55660.2022.10020785>.
- Silva Guerra, H. (2012). Panorama del negocio minorista en Colombia. *Pensamiento & Gestión*, 32, 115-141.
- Singh Yadav N. *et al.* (2022). Business Decision making using Data Science. *2022 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES)*, Chennai, India (pp. 1-11). <https://doi.org/10.1109/ICSES55317.2022.9914352>
- Toro Ocampo, E. M., Mejía Giraldo, D. A. & Salazar Isaza, H. (2004). Pronóstico de ventas usando redes neuronales. *Scientia et Technica*, X(26), 25-30.

- Van der Voort H., Van Bulderen, S., Cunningham, S. & Janssen, M. (2021). Data science as knowledge creation a framework for synergies between data analysts and domain professionals. *Technological Forecasting and Social Change*, 173, 121160. <https://www.sciencedirect.com/science/article/pii/S004016252100593X>
- Veres, O., Ilchuk, P. & Kots, O. (2021). Data Science Methods in Project Financing Involvement. *2021 IEEE 16th International Conference on Computer Sciences and Information Technologies (CSIT)*, Lviv, Ukraine, (pp. 411-414). <https://doi.org/10.1109/CSIT52700.2021.9648679>
- Wirth, R. & Hipp, J. (2000). CRISP-DM: Towards a standard process model for data mining. *Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining*, 1, 29-39.
- Wu, C. & Xiao, L. (2021). Evidence based on patient's experience data and clinical guidelines-for patient-oriented clinical decision support. *2021 International Conference on Public Health and Data Science (ICPHDS)*, Chengdu, China, (pp. 240-247). <https://doi.org/10.1109/ICPHDS53608.2021.00056>.