



UNIVERSIDAD NACIONAL DE CONCEPCIÓN
FACULTAD DE CIENCIAS ECONÓMICAS Y ADMINISTRATIVAS
CARRERA DE INGENIERIA EN INFORMATICA EMPRESARIAL



**ANALISIS PREDICTIVO DE DATOS DE NEGOCIOS DEL SECTOR RETAIL
UTILIZANDO ALGORITMOS DE CLUSTERING**

TESIS DE GRADO PARA OBTENER EL TÍTULO DE
INGENIERO/A EN INFORMÁTICA EMPRESARIAL

AUTOR:
ISIDRO DANIEL CHÁVEZ VARELA

AGOSTO 2023
CONCEPCIÓN – PARAGUAY



UNIVERSIDAD NACIONAL DE CONCEPCIÓN
FACULTAD DE CIENCIAS ECONÓMICAS Y ADMINISTRATIVAS



**ANALISIS PREDICTIVO DE DATOS DE NEGOCIOS DEL SECTOR RETAIL
UTILIZANDO ALGORITMOS DE CLUSTERING**

**TESIS DE GRADO PARA OBTENER EL TÍTULO DE
INGENIERO/A EN INFORMÁTICA EMPRESARIAL**

AUTOR:
ISIDRO DANIEL CHÁVEZ VARELA

DIRECTOR DE TESIS
LIC. DIEGO PEDRO PINTOS ROA

ABRIL 2023
CONCEPCIÓN – PARAGUAY

HOJA DE APROBACIÓN

ANALISIS PREDICTIVO DE DATOS DE NEGOCIOS DEL SECTOR RETAIL UTILIZANDO ALGORITMOS DE CLUSTERING

TESIS DE GRADO PRESENTADA PARA OBTENER EL TÍTULO DE
INGENIERO/A EN INFORMÁTICA EMPRESARIAL

AUTOR

ISIDRO DANIEL CHAVEZ VARELA

DIRECTOR DE TESIS: LIC. DIEGO PEDRO PINTOS ROA

TRIBUNAL DE EXPOSICIÓN Y DEFENSA DE LA TESIS

Resultado de la evaluación:

Isidro Daniel Chávez Varela:

Número Letra

DECLARACIÓN DE RESPONSABILIDAD

Isidro Daniel Chávez Varela

DECLARO QUE:

El trabajo de tesis de grado denominado “**ANALISIS PREDICTIVO DE DATOS DE NEGOCIOS DEL SECTOR RETAIL CON ALGORITMOS DE CLUSTERING**”, de la carrera de Ingeniería en Informática Empresarial, ha sido desarrollado con base en una investigación exhaustiva, respetando derechos intelectuales de terceros, conforme las citas bibliográficas cuyas fuentes se incorporan en la referencia bibliográfica.

Consecuentemente este trabajo es de mi autoría.

En virtud de esta declaración, me responsabilizo del contenido, veracidad y alcance del mismo.

Concedo a la FCEA un trabajo original y las copias a fin de que puedan servir de consultas en la Biblioteca de la Institución.

Isidro Daniel Chávez Varela

Concepción, 2023

DEDICATORIA

A mis padres: María y Máximo

Por apoyarme siempre y por el amor que
siempre me brindan.

AGRADECIMIENTOS

A mi familia por el apoyo de siempre,

A los docentes y funcionarios de la Facultad, quienes siempre me apoyaron y brindaron ayuda cuando necesitaba,

Al Profesor. Diego Pedro Pintos Roa por brindar su valiosa colaboración y orientación para la elaboración del trabajo,

A los amigos y compañeros de la carrera por alentarme a seguir adelante,

A todas las personas que no alcancó a nombrar, quienes me ayudaron de alguna manera.

ÍNDICE

RESUMEN	x
ABSTRACT	xii
INTRODUCCIÓN.....	14
CAPÍTULO I. PROBLEMA Y PROPÓSITO	16
Antecedentes del tema.....	17
Planteamiento del Problema	20
Preguntas de investigación	22
Pregunta General.....	22
Preguntas Específicas	22
Objetivos de la investigación	23
Objetivo General	23
Objetivos Específicos	23
Justificación	24
Delimitaciones y Límites.....	26
CAPÍTULO II. MARCO TEÓRICO	27
Inteligencia de Negocios	28
Descubrimiento del conocimiento en la base de Datos. (KDD, Knowledge Discovery in Databases)	28
Minería de Datos	29
Análisis de Técnicas de data mining	30
Análisis de agrupamiento o clustering.....	30
CAPÍTULO III. MATERIALES Y METODOS.....	35

En KNIME, el "Lector de Excel" es un nodo que te permite importar datos desde archivos Excel (.xlsx) al flujo de trabajo	42
Marco Legal	48
CAPÍTULO III. MARCO METODOLÓGICO.....	50
 Enfoque de la investigación	51
 Nivel de la investigación.....	51
 Diseño de la investigación.....	52
 Determinación de la población y selección de la muestra	53
 Población.....	53
 Muestra.....	54
 Muestreo.....	54
 Variable de estudio	55
 Técnica de recolección de datos	55
 Operacionalización de Variables.....	57
 Técnicas de procesamiento y análisis de datos	59
 CAPÍTULO IV. ANÁLISIS E INTERPRETACIÓN DE LOS RESULTADOS..	
 61	
 Resultados Obtenidos.....	62
 Dimensión 1. Cantidad de Clusters	62
 CAPÍTULO V. CONCLUSIONES Y RECOMENDACIONES.....	64
 Conclusiones	65
 Recomendaciones	67
 Referencias	68

RESUMEN

El presente estudio se enfocó en realizar un análisis predictivo de datos de negocios en el sector retail utilizando algoritmos de clustering. El objetivo principal fue examinar la efectividad y las ventajas de la aplicación de algoritmos de clustering en la predicción de tendencias y patrones de compra en establecimientos minoristas. Para lograr este propósito, se empleó una combinación de datos históricos de ventas, información de productos y perfiles de clientes.

En el proceso de investigación, se recopilaron y prepararon datos de ventas anteriores, datos de productos y detalles de los clientes. Además, se llevó a cabo una revisión detallada de la literatura existente sobre técnicas de análisis predictivo y algoritmos de clustering en el ámbito del sector retail.

Para validar la hipótesis de que la aplicación de algoritmos de clustering puede mejorar la precisión de las predicciones de ventas, se procedió a realizar los siguientes pasos:

Selección de Variables y Algoritmo de Clustering:

Se seleccionaron variables clave, como históricas de ventas, características del producto y comportamiento del cliente.

Se optó por el algoritmo de clustering k-means debido a su capacidad para identificar grupos similares en los datos.

Generación de Clusters y Modelado Predictivo:

Se ejecutó el algoritmo de clustering en los datos y se generaron clusters de productos y perfiles de clientes similares.

Los clusters obtenidos se utilizaron como características en un modelo de regresión para predecir las ventas futuras de productos específicos.

Validación y Evaluación del Modelo:

Se dividió el conjunto de datos en conjuntos de entrenamiento y prueba para validar el modelo predictivo.

Se evaluó el rendimiento del modelo utilizando métricas como el error cuadrado medio y el coeficiente de determinación.

Aplicación de Resultados y Toma de Decisiones:

Utilizando las predicciones generadas por el modelo, se identificaron productos con mayores probabilidades de ventas futuras.

Estos resultados se utilizaron para informar las decisiones estratégicas, como la planificación de inventario y la optimización de las estrategias de marketing.

Los resultados del análisis demostraron que la aplicación de algoritmos de clustering en la predicción de ventas en el sector retail puede mejorar significativamente la precisión de las predicciones. Esto permite a los minoristas tomar decisiones más informadas y estratégicas para optimizar sus operaciones y maximizar sus ganancias.

En resumen, este estudio destacó la importancia y la efectividad de utilizar técnicas de análisis predictivo con algoritmos de clustering en el sector retail. Los resultados obtenidos respaldaron la hipótesis planteada y resaltaron el potencial de estas técnicas para impulsar la toma de decisiones inteligentes en el ámbito empresarial.

Palabras Claves: Algoritmos, Clustering; Retail.

ABSTRACT

The present study focused on conducting predictive analysis of business data in the retail sector using clustering algorithms. The main objective was to examine the effectiveness and advantages of applying clustering algorithms in predicting trends and purchase patterns in retail establishments. To achieve this purpose, a combination of historical sales data, product information, and customer profiles was employed.

In the research process, data from previous sales, product data, and customer details were collected and prepared. Additionally, a thorough review of existing literature on predictive analysis techniques and clustering algorithms in the context of the retail sector was conducted.

To validate the hypothesis that the application of clustering algorithms can enhance the accuracy of sales predictions, the following steps were taken:

Variable Selection and Clustering Algorithm:

Key variables were selected, such as sales history, product features, and customer behavior.

The k-means clustering algorithm was chosen due to its ability to identify similar groups within data.

Cluster Generation and Predictive Modeling:

The clustering algorithm was executed on the data, generating clusters of similar products and customer profiles.

The obtained clusters were used as features in a regression model to predict future sales of specific products.

Model Validation and Evaluation:

The dataset was divided into training and testing sets to validate the predictive model.

The model's performance was evaluated using metrics such as mean squared error and coefficient of determination.

Application of Results and Decision Making:

Using predictions generated by the model, products with higher probabilities of future sales were identified.

These results were utilized to inform strategic decisions, such as inventory planning and optimization of marketing strategies.

The analysis results demonstrated that applying clustering algorithms in sales prediction

within the retail sector can significantly improve prediction accuracy. This enables retailers to make more informed and strategic decisions to optimize their operations and maximize profits.

In summary, this study highlighted the importance and effectiveness of employing predictive analysis techniques with clustering algorithms in the retail sector. The obtained results supported the proposed hypothesis and underscored the potential of these techniques in driving intelligent decision-making in the business realm.

Keywords: Algorithms, Clustering; Retail.

INTRODUCCIÓN

El análisis predictivo de datos de negocios en el sector retail mediante algoritmos de clustering está emergiendo como una poderosa herramienta para informar decisiones estratégicas y operativas en la industria minorista. Con el avance continuo de la tecnología y la creciente disponibilidad de datos, las empresas están recurriendo cada vez más a técnicas analíticas avanzadas para comprender mejor los patrones de comportamiento de los clientes y anticipar las tendencias del mercado.

Este enfoque se basa en la premisa de que los datos generados por las transacciones, el comportamiento de los clientes y otros factores relevantes en el sector retail contienen información valiosa que puede ser explotada para tomar decisiones informadas y maximizar la eficiencia operativa. Los algoritmos de clustering, en particular, permiten agrupar conjuntos de datos similares y encontrar patrones ocultos que podrían no ser evidentes a simple vista.

En esta investigación, el objetivo principal es aplicar el análisis predictivo de datos de negocios utilizando algoritmos de clustering en el sector retail. Esto implica no solo recopilar y analizar datos históricos de ventas y comportamiento del cliente, sino también utilizar estos datos para predecir patrones futuros. Mediante la identificación de grupos o clusters de clientes con comportamientos de compra similares, las empresas pueden personalizar sus estrategias de marketing y promoción para cada grupo, mejorando así la experiencia del cliente y aumentando las tasas de conversión.

Este enfoque metodológico se caracteriza por su naturaleza cuantitativa y su alcance descriptivo. Se busca describir y analizar los datos de ventas, comportamiento del cliente y otros factores relevantes en el sector retail, a fin de identificar patrones y tendencias que

puedan ser utilizados para predecir futuras decisiones de compra. Al utilizar algoritmos de clustering, se espera identificar grupos de clientes con similitudes en sus preferencias y comportamientos, lo que permitirá a las empresas diseñar estrategias de segmentación más efectivas y mejorar la toma de decisiones en áreas como el inventario y la planificación de productos.

En resumen, el análisis predictivo de datos de negocios en el sector retail utilizando algoritmos de clustering representa una valiosa oportunidad para las empresas minoristas. Al aprovechar la riqueza de los datos disponibles y aplicar técnicas analíticas avanzadas, las organizaciones pueden anticipar las necesidades y preferencias de los clientes, lo que a su vez conduce a una toma de decisiones más estratégica y a la mejora general de la eficiencia operativa en el competitivo mundo del retail.

CAPÍTULO I. PROBLEMA Y PROPÓSITO

Antecedentes del tema

A continuación, se señalarán algunos estudios vinculados con el análisis de datos de negocios, las investigaciones presentadas estarán en ordenadas cronológicamente

Según Kusrini (2015), en su estudio titulado “Grouping of Retail Items by Using K-Means Clustering”, realizado en Yogyakarta, Indonesia, se presenta un método que utiliza clustering K-means. Concluyen que el clustering K-Means se puede emplear en el proceso de categorización de artículos en categorías de movimiento rápido y lento. Utilizando datos de ventas de Citramart Minimarket de STMIK AMIKOM Yogyakarta para los años 2013 y 2014, se demuestra que el mejor clúster se logra a través del proceso de agrupación con datos anuales y un valor de transacción variable. El valor del índice Xie-Beny para este clúster es 36,265.

Sokol y Cerny (2015) en su estudio titulado “Clustering retail products based on customer behaviour”, realizado en Praga, República Checa, encontraron que la categorización de productos minoristas es fundamental para la toma de decisiones empresariales. En este documento, se emplea un enfoque puramente basado en datos. La agrupación de productos se basa exclusivamente en el comportamiento del cliente. Proponen un método para agrupar productos minoristas utilizando datos de la cesta de mercado. Su modelo se formula como un problema de optimización resuelto mediante un algoritmo genético. Se demuestra en datos simulados cómo se comporta su método en diferentes entornos. La aplicación con datos reales de una empresa farmacéutica checa muestra resultados similares a la clasificación de expertos. El número de clústeres es un parámetro en su algoritmo, y se demuestra que permitir más clústeres que las categorías originales proporciona información adicional sobre la estructura de la categorización de productos.

En el trabajo de Chen, Fang, Yang, Nie, Zhao y Zhexue (2018), denominado "PurTreeClust: A Clustering Algorithm for Customer Segmentation from Massive Customer Transaction Data", se destaca que la agrupación de datos de transacciones de clientes es fundamental para analizar el comportamiento del cliente en las empresas minoristas y de comercio electrónico. Los productos de las empresas a menudo se organizan en un árbol de productos, donde los nodos hoja son productos para la venta y los nodos internos (excepto el nodo raíz) pueden representar diversas categorías de productos. Proponen el concepto de "árbol de compras personalizado", denominado árbol de compras, para representar los registros de transacciones de un cliente. Por lo tanto, el conjunto de datos de transacciones de clientes se puede comprimir en un conjunto de árboles de compras. Proponen un algoritmo de clustering particional llamado PurTreeClust para la agrupación rápida de árboles de compras. Introducen una nueva métrica de distancia para calcular de manera efectiva la distancia entre dos árboles de compras. Para agrupar los datos de árboles de compras, primero clasifican los árboles de compras como árboles representativos candidatos con un nuevo criterio de densidad separada, y luego seleccionan los principales k clientes como representantes de k grupos de clientes. Finalmente, los resultados de la agrupación se obtienen asignando a cada cliente al representante más cercano. También proponen un método basado en estadísticas de brechas para evaluar el número de clústeres. Se realizan una serie de experimentos en diez conjuntos de datos de transacciones de la vida real, y los resultados experimentales demuestran el rendimiento superior del método propuesto.

Según Yoseph y Heikkila (2018), en su estudio titulado "Segmenting Retail Customers with an Enhanced RFM and a Hybrid Regression/Clustering Method", realizado en Turku, Finlandia, se menciona que las estrategias de marketing dirigidas a menudo pasan por alto la evolución del comportamiento del cliente a lo largo del tiempo, lo que puede llevar a concentrarse en clientes no rentables. Combinan las puntuaciones RFM con un modelo de valor de vida útil del cliente para segmentar a los clientes de una tienda minorista de moda y accesorios de tamaño mediano en Kuwait. Un algoritmo de regresión

modificado investiga el comportamiento de compra del cliente, extrayendo información de los datos del punto de venta para guiar la toma de decisiones informada. La agrupación se lleva a cabo mediante los métodos K-means y Expectation Maximization. El análisis de la calidad de los clústeres revela que el primero supera al segundo en la identificación de segmentos de mercado relevantes y estrategias de marketing apropiadas.

Planteamiento del Problema

En el dinámico entorno del sector retail, la toma de decisiones informadas se ha convertido en un factor crítico para el éxito y la supervivencia de las organizaciones. La creciente disponibilidad de datos y el avance tecnológico han brindado nuevas oportunidades para comprender las preferencias cambiantes de los consumidores y anticipar las tendencias del mercado. En este contexto, el análisis predictivo de datos y el uso de algoritmos de clustering se presentan como herramientas potencialmente transformadoras para la industria minorista.

Sin embargo, a pesar de las posibilidades que ofrece el análisis predictivo de datos y los algoritmos de clustering, existen desafíos significativos que deben abordarse. En primer lugar, el sector retail opera en un entorno altamente competitivo y sensible al tiempo, donde las decisiones tardías o imprecisas pueden resultar en pérdida de clientes y oportunidades de ventas. Además, el proceso de adopción y aplicación de estas técnicas analíticas en el ámbito del retail puede ser complejo debido a la necesidad de comprender, adaptar y aplicar los resultados de manera efectiva.

Además, se debe considerar cómo el análisis predictivo y los algoritmos de clustering pueden ser implementados de manera eficiente en una amplia variedad de contextos minoristas, cada uno con sus propias características y desafíos únicos. La capacidad de generalizar y adaptar estos enfoques a diferentes tipos de tiendas, productos y segmentos de clientes es un aspecto crítico que requiere una cuidadosa consideración.

En este sentido, el presente estudio busca abordar estos desafíos identificados y responder a la siguiente pregunta de investigación: ¿Cómo puede el análisis predictivo de datos de negocios en el sector retail, mediante el uso de algoritmos de clustering, optimizar la toma de decisiones estratégicas y operativas para mejorar la eficiencia y el rendimiento empresarial?

Responder a esta pregunta permitirá identificar los beneficios específicos y las limitaciones potenciales de la aplicación de técnicas de análisis predictivo y algoritmos de clustering en el contexto del sector retail. Asimismo, contribuirá a comprender cómo estas herramientas pueden ser adaptadas y aprovechadas para abordar los desafíos inherentes a la toma de decisiones en un entorno altamente competitivo y en constante evolución.

En resumen, el presente estudio se enfoca en explorar cómo el análisis predictivo de datos y los algoritmos de clustering pueden ser aprovechados para optimizar la toma de decisiones en el sector retail. Mediante la identificación y comprensión de los desafíos actuales en esta área, se busca proporcionar recomendaciones y directrices prácticas para la aplicación efectiva de estas técnicas, contribuyendo así al avance y la mejora de la industria minorista en un contexto globalmente competitivo y tecnológicamente impulsado.

Preguntas de investigación

Pregunta General

¿Cuales son los resultados que se han logrado al implementar la metodología de análisis predictivo a través de la herramienta KNIME en conjuntos de datos de negocios dentro del sector retail?

Preguntas Específicas

¿Cuáles son las variaciones en los patrones de comportamiento de los clientes en el sector retail al implementar algoritmos de clustering mediante la herramienta KNIME?

¿Cuáles son las ventajas específicas que se han derivado de la aplicación de la metodología de análisis predictivo en la optimización de estrategias comerciales en el sector retail?

¿En qué medida ha influido la implementación de la metodología de análisis predictivo utilizando KNIME en la identificación de patrones de comportamiento del consumidor en el sector retail?

Objetivos de la investigación

Objetivo General

Analizar los resultados que se han logrado al implementar la metodología de análisis predictivo a través de la herramienta KNIME en conjuntos de datos de negocios dentro del sector retail.

Objetivos Específicos

Describir las variaciones en los patrones de comportamiento de los clientes en el sector retail al implementar algoritmos de clustering mediante la herramienta KNIME.

Demostrar las ventajas específicas que se han derivado de la aplicación de la metodología de análisis predictivo en la optimización de estrategias comerciales en el sector retail.

Evaluuar el impacto de la implementación de la metodología de análisis predictivo utilizando KNIME en la identificación de los patrones de comportamiento del consumidor en el sector retail.

Justificación

La presente investigación se erige con significativas implicaciones teóricas, prácticas, sociales y metodológicas, además de ser altamente viable en su ejecución.

Desde una perspectiva teórica, este estudio se erige como una fuente valiosa de conocimiento en el campo del análisis predictivo de datos de negocios en el sector retail. Los aportes de diversos autores que respaldan la aplicación de algoritmos de clustering mediante KNIME fortalecen el andamiaje de variables y guían la trayectoria investigativa. El análisis teórico demuestra que la conjunción de la metodología de análisis predictivo con la herramienta KNIME permite desentrañar patrones clave en el comportamiento del consumidor, esencial en el entorno retail.

La implicación práctica de esta investigación radica en la potencialidad de sus resultados para fomentar estrategias concretas en la optimización de estrategias comerciales en el sector retail. La aplicación de análisis predictivo y algoritmos de clustering mediante KNIME confiere la capacidad de segmentar con precisión el mercado y personalizar las tácticas empresariales. Este enfoque revoluciona la toma de decisiones, permitiendo una alineación más cercana con las preferencias cambiantes de los consumidores y una eficiencia operativa elevada.

Desde un punto de vista social, esta investigación contribuye a los beneficios que aportan las innovaciones tecnológicas en el sector retail. La mejora en la comprensión del comportamiento del consumidor y la personalización de estrategias impulsan la calidad de vida de los ciudadanos al asegurar experiencias de compra más satisfactorias y relevantes. La aplicación de análisis predictivo y algoritmos de clustering contribuye a la seguridad y la eficiencia del mercado, evitando retrasos y optimizando la circulación de productos y servicios.

La trascendencia metodológica de este estudio radica en su rigor científico, lo que lo convierte en un referente valioso para futuras investigaciones en el campo de análisis de datos y estrategias empresariales. La aplicación de la metodología de análisis predictivo mediante KNIME en datos de negocios del sector retail sienta las bases para un análisis más profundo y más informado en futuros trabajos de investigación.

En resumen, esta investigación se erige como una contribución sólida y esencial al campo del análisis de datos aplicado al sector retail, enriqueciendo tanto la teoría como la práctica en el proceso. Los resultados obtenidos brindarán perspectivas claras y oportunas para la toma de decisiones informadas en las empresas del sector, fortaleciendo su posición competitiva en un mercado en constante evolución.

Delimitaciones y Límites

La presente investigación se ha acotado al análisis de datos de negocios dentro del ámbito del sector retail, empleando la metodología de análisis predictivo mediante algoritmos de clustering con la herramienta KNIME. Esta delimitación se establece en base a las siguientes consideraciones:

Ámbito de Estudio: La investigación se enfoca exclusivamente en el análisis de datos de negocios pertenecientes al sector retail, limitándose a datos disponibles públicamente y de carácter anónimo para su análisis.

Limitaciones: Entre las limitaciones, se puede mencionar la potencial falta de actualización de registros relacionados con los datos de negocios en el sector retail. Esto podría tener un impacto en la precisión de los resultados obtenidos al identificar patrones y tendencias en los datos.

El área de investigación se sitúa en el campo de la inteligencia de negocios y descubrimiento de conocimiento en la base de datos.

CAPÍTULO II. MARCO TEÓRICO

Inteligencia de Negocios

La inteligencia de negocios (BI, por sus siglas en inglés) es una disciplina estratégica que involucra la recopilación, procesamiento y presentación de datos relevantes para la toma de decisiones empresariales informadas. En el contexto del sector retail, la inteligencia de negocios se convierte en un medio esencial para comprender el comportamiento del consumidor, identificar oportunidades y mejorar la eficiencia operativa (Turban et al., 2011).

Descubrimiento del conocimiento en la base de Datos. (KDD, Knowledge Discovery in Databases)

El proceso de Descubrimiento de Conocimiento en Bases de Datos (KDD, por sus siglas en inglés) representa una fusión entre descubrimiento y análisis automatizado. Este proceso implica la extracción de patrones significativos, como reglas o funciones, a partir de datos almacenados, con el propósito de que puedan ser analizados por los usuarios (Agrawal y Srikant, 1994) (Chen, Han y Yu, 1996) (Piatetsky Shapiro, Brachman y Khabaza, 1996) (Han y Kamber, 2001).

El proceso consta de una secuencia iterativa como se detalla a continuación:

- 1) Selección de datos: los datos relevantes para la orden del analista se recuperan de la base de datos.

- 2) Pre-procesamiento de datos: Consiste en preparar los datos eliminando ruido y datos inconsistentes. Es la fase más laboriosa para que los datos sean fiables.
- 3) Transformación de datos: los datos se transforman o consolidan en formas apropiadas para la minería por formando operaciones de resumen o agregación.
- 4) Minería de datos: el proceso esencial para extraer patrones de datos.
- 5) Evaluación y presentación de patrones: identificar los patrones interesantes que representan el conocimiento, basado en algunas medidas de interés, y las técnicas de visualización y representación del conocimiento son utilizados para presentar el conocimiento extraído al usuario.

Minería de Datos

La minería de datos es una parte integral de la inteligencia de negocios, centrada en el descubrimiento de patrones y relaciones significativas en grandes conjuntos de datos. Este proceso implica la utilización de algoritmos y técnicas avanzadas para extraer información valiosa de los datos almacenados. En el contexto del sector retail, la minería de datos permite el análisis de los patrones de compra, segmentación de clientes y pronóstico de tendencias (Han, Kamber, & Pei, 2011).

Estos datos son muy necesarios para extraer conclusiones e información relevante que se necesita obtener para determinadas situaciones. Gracias a la minería de datos se puede predecir escenarios futuros como uno de los factores que posee, se le denominó análisis predictivo. (Microsoft, 2016).

Análisis de Técnicas de data mining

Las técnicas de data mining sirven para los procesos de decisión dentro de un negocio, el objetivo es descubrir hechos a través de la recopilación de datos que ayuden a mejorar el presente y futuro de la empresa y obtener las respuestas necesarias para entender por qué suceden estos hechos.

Análisis de agrupamiento o clustering

El Análisis de agrupamiento, conocido como Análisis de conglomerados, es una técnica estadística multivariada cuyo propósito es agrupar un conjunto de objetos, tratando de lograr la máxima homogeneidad en cada grupo y la mayor diferencia entre los grupos.

Los Reglas de Asociacion

Las reglas de asociación son una técnica esencial en la minería de datos que identifica patrones de comportamiento en los datos. En el sector retail, estas reglas pueden revelar relaciones entre productos y preferencias de compra, permitiendo la personalización de ofertas y estrategias (Agrawal & Srikant, 1994).

Algoritmos de Clustering

Los algoritmos de clustering agrupan datos similares en clústeres con características comunes. En el contexto del sector retail, estos algoritmos permiten segmentar a los consumidores en grupos homogéneos con patrones de compra similares. Esto facilita la personalización de estrategias y ofertas, así como la identificación de tendencias emergentes (Han et al., 2011).

Algoritmos de K-Means

El algoritmo K-Means es una técnica de clustering que agrupa datos en k clústeres, siendo k un valor predefinido. Esta técnica es eficaz para la segmentación de datos en grupos homogéneos. En el contexto del análisis de datos de negocios en el sector retail, el algoritmo K-Means es utilizado para identificar patrones de compra similares entre los consumidores (Hartigan & Wong, 1979).

Análisis Predictivo

El análisis predictivo es una disciplina que utiliza datos históricos y técnicas estadísticas para prever tendencias y comportamientos futuros. En el sector retail, el análisis predictivo se aplica para anticipar patrones de compra, demanda de productos y preferencias de los consumidores. Esta información permite la toma de decisiones proactivas y la adaptación de estrategias (Witten & Frank, 2005)..

Patrones de Compra

Los patrones de compra representan comportamientos recurrentes de los consumidores al interactuar con productos y servicios en el sector retail. Estos patrones reflejan las preferencias, frecuencia y tipos de productos adquiridos. La identificación de patrones de compra es crucial para segmentar a los consumidores y adaptar estrategias de marketing de manera efectiva.

Los Métodos de agrupamiento

En la literatura existen una variedad de métodos y algoritmos para la agrupación de objetos. La selección de un algoritmo de agrupamiento, depende generalmente de la cantidad y tipo de datos disponibles y el propósito de su aplicación. Una categorización de los algoritmos de agrupamiento son los siguientes:

1. **Métodos jerárquicos:** Se basan en un proceso secuencial para la formación de los grupos o clúster, a través de establecer jerarquías entre los objetos o individuos. Pueden ser métodos aglomerativos o divisivos. **Método aglomerativo:** Se inicia con un número de clúster igual al número total de objetos y se van uniendo en forma aglomerativa de acuerdo a una métrica de distancias; al final se forma un sólo clúster con todos los objetos. Las técnicas para la formación de los clúster son: Enlace simple (vecino más cercano), Enlace completo (vecino más alejado), Enlace promedio, Enlace centroide,

Enlace Ward. Método de divisiones: Su proceso secuencial de formación de clúster es contrario al aglomerativo. Se inicia con un solo grupo o clúster, que contiene el total de objetos. Luego se van dividiendo (descendente) en subgrupos considerando los más alejados, hasta llegar a “n” grupos de un solo objeto. El más usado es: Algoritmo de Howard-Harris.

2.

2. Métodos basados en particiones: Es un método de clúster no jerárquico, donde el número de clúster a formarse k es un valor (parámetro) conocido. Los k grupos o clúster se construyen por un proceso iterativo de conformación de k particiones. El método consiste en distribuir (moviendo) los objetos en los k grupos, minimizando las distancias entre los objetos dentro del grupo con respecto a su centroide (la media, mediana, moda, medoides, etc.). Se inicia seleccionando k objetos aleatoriamente como los centroides iniciales para cada clúster. En cada iteración, se asigna el objeto al clúster más similar (la menor distancia con respecto al centroide) y se calcula el nuevo centroide de cada clúster y se termina con la asignación de todos los objetos bajo un criterio de parada de convergencia. Los algoritmos basados en particiones son: k-medias, k-medianas, k-medoides (PAM y CLARA), etc.

3. Métodos basados en modelo: Se basan en hallar un modelo para cada clúster, que mejor se ajuste los datos a cada clúster. El método COBWEB pertenece a los métodos de aprendizaje conceptual o basado en modelos. Esto significa que cada clúster se considera como un modelo descriptivo. El COBWEB, es considerado un cluster jerárquico y su aprendizaje se representa por un árbol de clasificación (árbol conceptual jerárquico), donde cada nodo es un Cluster (concepto) que tiene una descripción probabilística del concepto que resume los objetos clasificados en él. El algoritmo se aplica atributos cualitativos (COBWEB) y se extiende a numéricos usando la Normal (CLASSIST). COBWEB depende del orden de los objetos (se recomienda probar con los objetos en diferente orden).

4. Métodos probabilísticos: Se basan en hallar y usar funciones de densidades como medida de aproximación. El EM (Expectation Maximization), es conocido como el clustering probabilístico. El EM, busca el grupo de clústeres más probable dado un conjunto de ejemplos. Los ejemplos tienen cierta probabilidad de pertenecer a un grupo o cluster. Este clustering se basa en el modelo estadístico de mezcla de distribuciones. El EM, trata de obtener la función de densidad de probabilidades (FDP) desconocida para el conjunto de datos, haciendo una aproximación por una combinación lineal de las k distribuciones asociadas a cada cluster (mezcla de k distribuciones de probabilidades). La mezcla más sencilla se tiene cuando los atributos son numéricos con distribuciones gaussianas, determinándose k distribuciones normales con medias y variancias diferentes para cada cluster.

CAPÍTULO III. MATERIALES Y METODOS

3.1. Área de Estudio

El área de estudio se centra en una empresa del sector retail que tiene una trayectoria de 7 años en el mercado y ha establecido su reputación en la ciudad. La empresa comenzó su incursión en el mundo digital en 2014, lo que permitió la recopilación de una gran cantidad de datos. Para este estudio, se consideraron las transacciones realizadas a partir del año 2018, ya que en ese año la empresa amplió su catálogo de productos.

3.2. Colección de Datos

Para obtener los datos necesarios, se realizó una copia de la base de datos de la empresa retail en un entorno de servidor de base de datos. La empresa proporcionó el backup de las transacciones realizadas hasta diciembre de 2020, por lo que los datos del año 2021 y 2022 no se incluyen en el análisis.

Se implementó un script SQL previamente desarrollado en investigaciones anteriores para obtener los registros de ventas de productos. Estos datos ya habían sido preprocesados y estaban listos para ser utilizados en la fase de Minería de Datos.

3.3. Preparación y Transformación del Conjunto de Datos para el Análisis de Minería de Datos

Para llevar a cabo el análisis de minería de datos, fue necesario formatear el conjunto de datos obtenido de la consulta SQL. La herramienta utilizada para el análisis requería que los datos estuvieran en un formato específico para registros de ventas, donde cada registro correspondía a una transacción de venta de productos. Las columnas representaban los productos disponibles en la tienda.

Se exploraron varias soluciones para transformar los datos en el formato necesario. Finalmente, se encontró una solución que implicaba la utilización de funciones de manipulación de datos en el servidor de base de datos. Se llevó a cabo una serie de transformaciones para obtener el formato requerido para el análisis de minería de datos. Este proceso implicó una inversión considerable de tiempo en investigación y experimentación para garantizar que los datos estuvieran en el formato correcto para el análisis de minería de datos. Uno de los productos más destacados y vendidos por la empresa es el gas, el cual se incluyó en el conjunto de datos para el análisis.

Gráfico N° 1. Datos importados en la herramienta KNIME.

bles

[Table](#) [Statistics](#)

fecha String	anulado Number (integer)	saldo Number (integer)	fecha_time Local Date Time	item Number (integer)	idproducto Number (integer)	producto String	cantidad Number (double)
2018-05-18 00:00:00	0	0	2018-05-18T00:00:00	1	111	AZUCAR BR	1
2018-05-18 00:00:00	0	0	2018-05-18T00:00:00	2	7	CAFE BELEN	1
2018-05-18 00:00:00	0	0	2018-05-18T00:00:00	3	75	BOLSA P HAMB...	2
2018-05-18 00:00:00	0	0	2018-05-18T00:00:00	4	1	GAS PETRONOR	1
2018-05-18 00:00:00	0	0	2018-05-18T00:00:00	1	1	GAS PETRONOR	2
2018-05-18 00:00:00	0	0	2018-05-18T00:00:00	2	54	CARAMELO MO...	3
2018-05-18 00:00:00	0	0	2018-05-18T00:00:00	3	1	GAS PETRONOR	1
2018-05-18 00:00:00	0	0	2018-05-18T00:00:00	4	1	GAS PETRONOR	1
2018-05-18 00:00:00	0	0	2018-05-18T00:00:00	5	111	AZUCAR BR	1
2018-05-18 00:00:00	0	0	2018-05-18T00:00:00	6	1	GAS PETRONOR	1
2018-05-18 00:00:00	0	0	2018-05-18T00:00:00	1	37	PAPEL HIGIENI...	2
2018-05-18 00:00:00	0	0	2018-05-18T00:00:00	2	1	GAS PETRONOR	1
2018-05-18 00:00:00	0	0	2018-05-18T00:00:00	1	100	GASEOSA PEPSI	6
2018-05-18 00:00:00	0	0	2018-05-18T00:00:00	2	101	GASEOSA MIRI...	6
2018-05-18 00:00:00	0	0	2018-05-18T00:00:00	3	61	SALSA DE SOJA...	1
2018-05-18 00:00:00	0	0	2018-05-18T00:00:00	4	34	SALSA PIMIENT...	2
2018-05-18 00:00:00	0	0	2018-05-18T00:00:00	5	46	SERVILLETAS D...	1
2018-05-18 00:00:00	0	0	2018-05-18T00:00:00	6	128	KETCHUP SUN	10
2018-05-18 00:00:00	0	0	2018-05-18T00:00:00	7	50	PALITOS P ASA...	2

Para asegurar la calidad de los datos y facilitar el análisis, se llevaron a cabo varios pasos de preparación y transformación en el conjunto de datos obtenido. En primer lugar, se eliminaron los campos vacíos o registros incompletos, lo que contribuyó a la integridad y consistencia de los datos.

Además, como parte del proceso de preparación de datos, se realizó una omisión estratégica. Uno de los productos más destacados y vendidos por la empresa es el gas, el cual se omitió temporalmente del conjunto de datos con el propósito de analizar cómo se comportaban las demás ventas sin su influencia. Esto permitió obtener una visión más clara de los patrones y tendencias de compra de otros productos, lo que podría haber sido enmascarado por la alta frecuencia de ventas de gas.

3.4. Introducción al algoritmo K-Means

El algoritmo K-Means es uno de los métodos de agrupación (clustering) más populares en el campo de la minería de datos y el aprendizaje automático. Su objetivo principal es agrupar un conjunto de datos en grupos o clústeres basándose en similitudes entre los puntos de datos. Aquí tienes más información sobre este algoritmo:

Inicialización:

1- Selección de Centroides Iniciales: Para K centroides iniciales, generalmente seleccionados de forma aleatoria entre los puntos de datos en el conjunto de datos. Puedes representar cada centroide como C_i .

Asignación:

2. Asignación de Puntos a Clústeres: Para cada punto de datos X_i , calculamos la distancia euclíadiana entre X_i y cada centroide C_j :

$$d(X_i, C_j) = \sqrt{\sum_{k=1}^n (X_{ik} - C_{jk})^2}$$

3. Reasignación y Actualización: Repetimos los pasos de asignación y recálculo hasta que se cumpla un criterio de detención, como la convergencia (cuando los centroides dejan de cambiar significativamente) o después de un número máximo de iteraciones.

Este proceso se repite hasta que los centroides convergen y los puntos ya no cambian de clúster en su mayoría. El resultado es una agrupación de datos en K clústeres basados en similitudes.

Es importante tener en cuenta que el rendimiento del algoritmo K-Means puede depender de la inicialización de los centroides y del valor de K, por lo que se suelen realizar múltiples ejecuciones con diferentes inicializaciones para evitar soluciones subóptimas.

Gráfico N° 2. Flujo de trabajo en la herramienta KNIME.

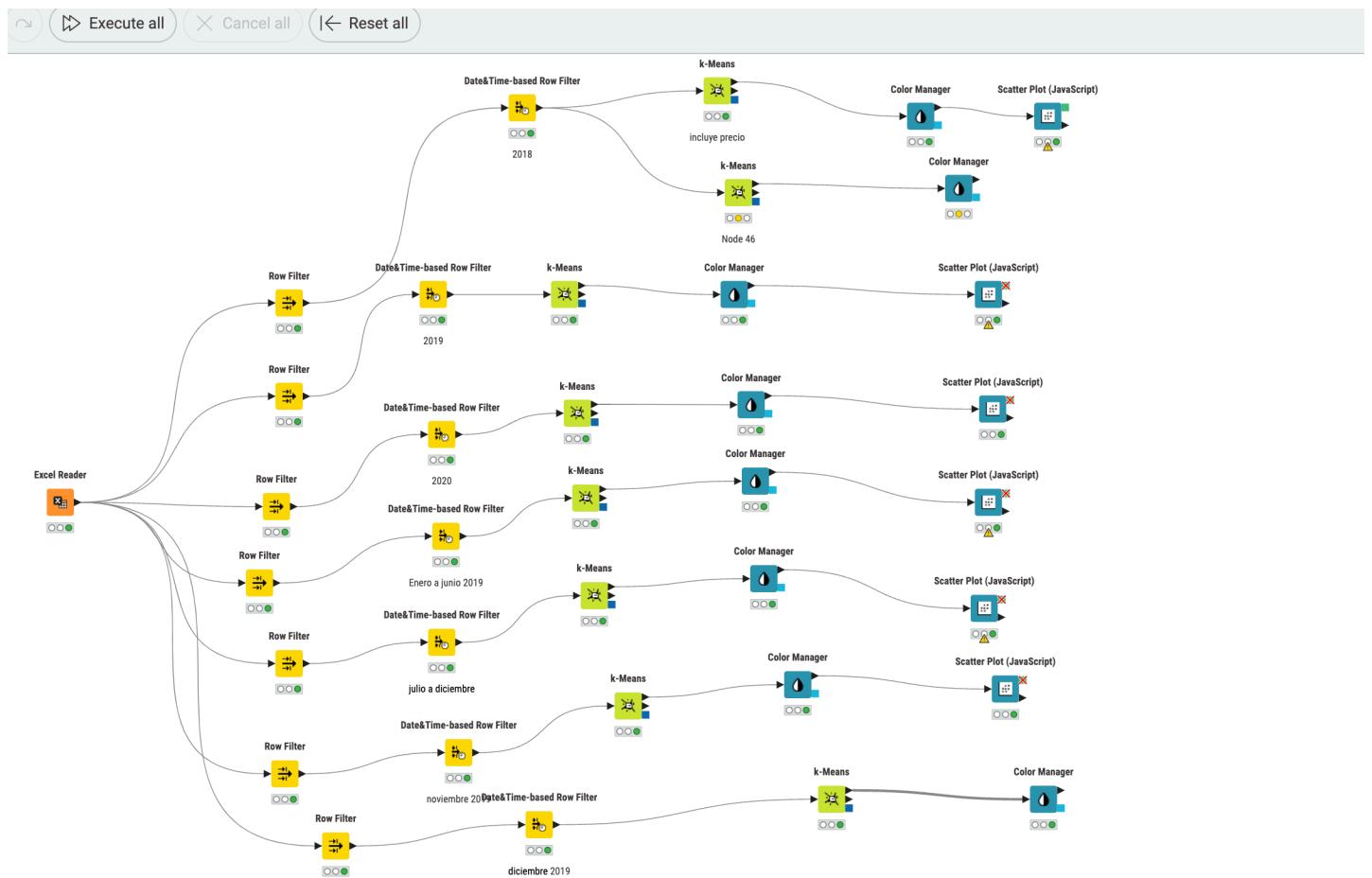


Gráfico N° 3. Nodo Lector De Excel.

En KNIME, el "Lector de Excel" es un nodo que te permite importar datos desde archivos Excel (.xlsx) al flujo de trabajo

Gráfico N° 4. Nodos de filtrado de filas, organización de fecha y hora.

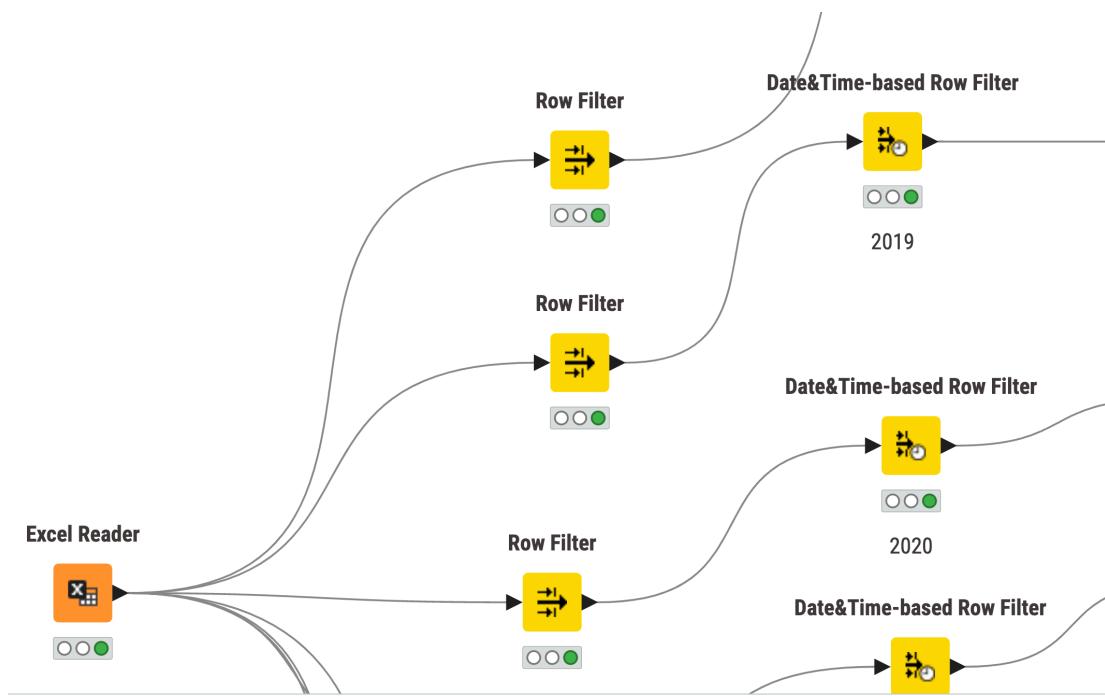


Gráfico N° 5. Configuración del nodo del algoritmo K-Means para asignar números de clusters y centroides

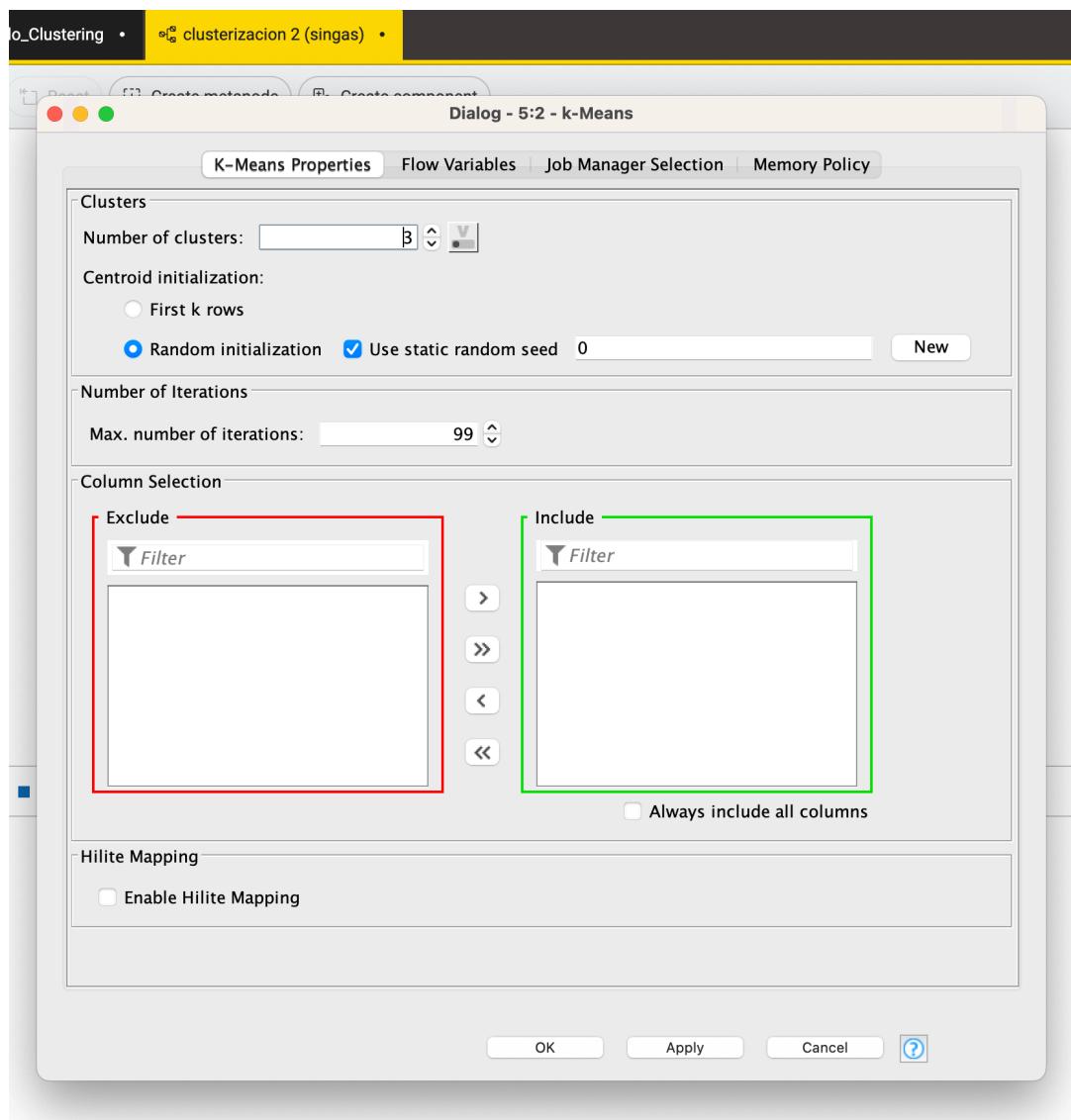


Gráfico N° 6. Configuración del nodo “Color manager” donde se asignan colores a los diferentes clusters

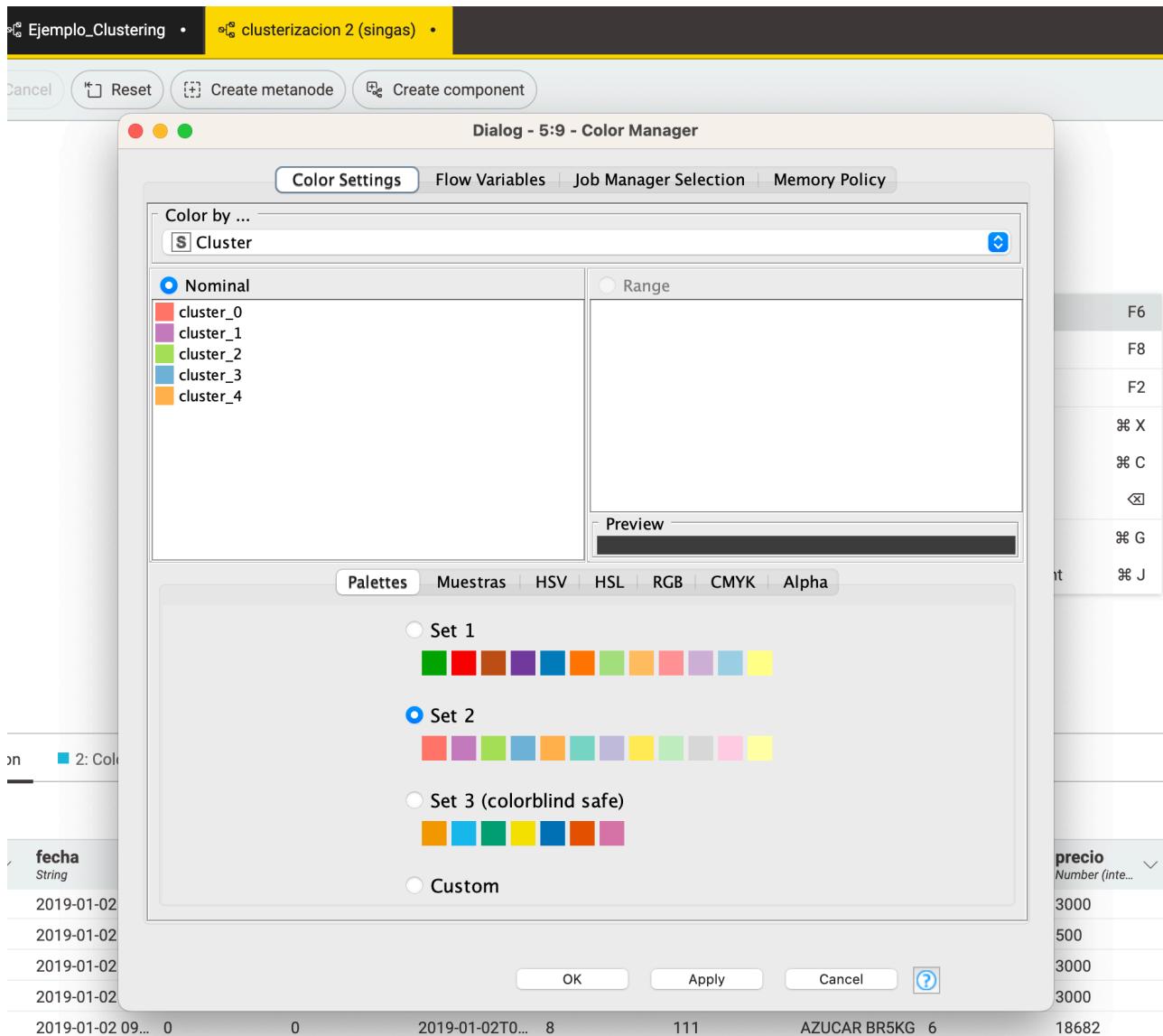
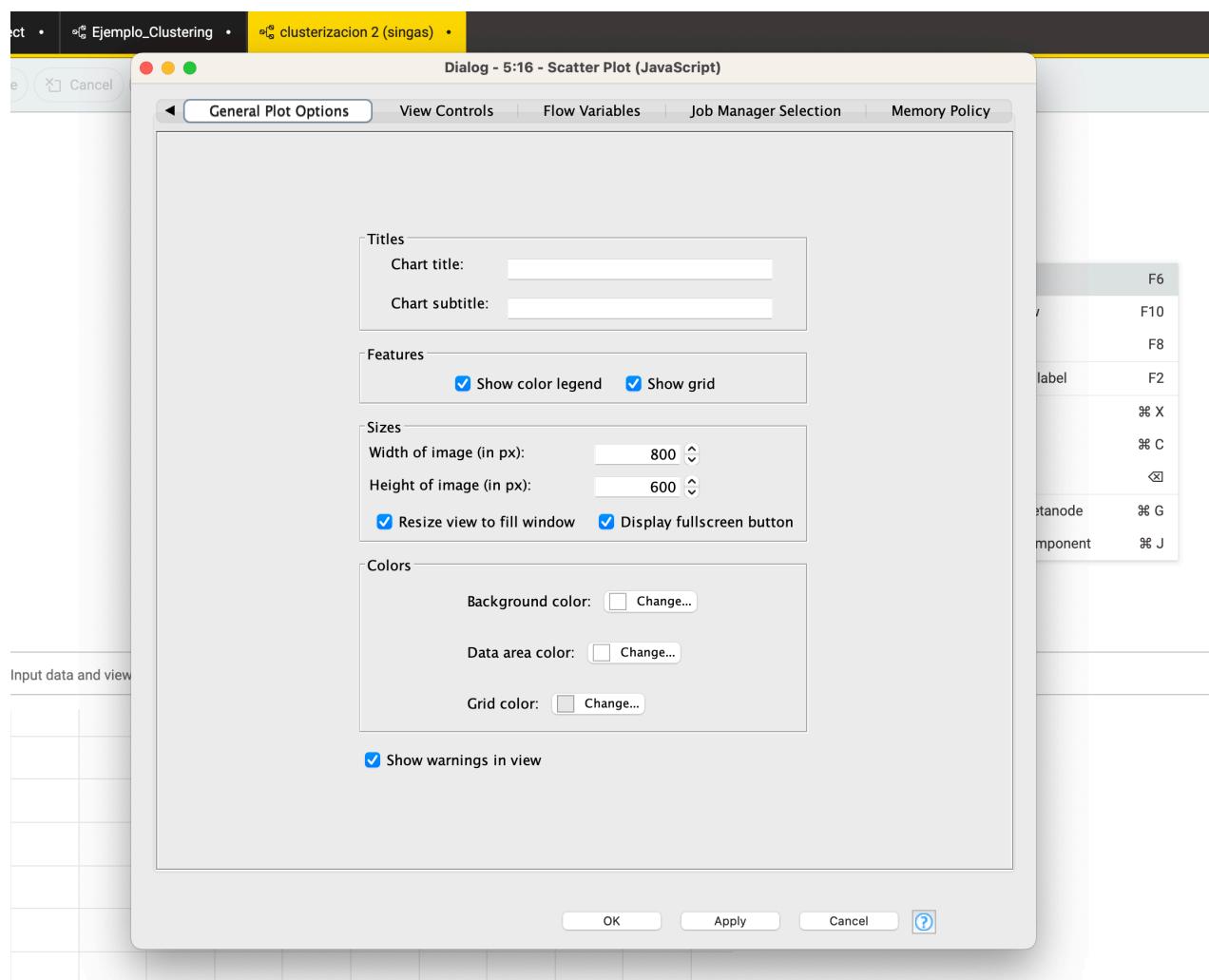


Gráfico N° 7. Configuración del nodo “Scarlet Plot” donde se configura como se van a visualizar los datos y particiones de clusters.



Marco Legal

El marco legal para esta investigación se fundamenta en las regulaciones y normativas relacionadas con la recopilación y el uso de datos en el ámbito empresarial y de investigación. Algunos de los aspectos legales clave a considerar incluyen:

Protección de Datos Personales: La recolección y el análisis de datos pueden involucrar información personal o confidencial. En este sentido, se debe cumplir con las leyes de protección de datos aplicables para garantizar la privacidad y la seguridad de los individuos cuyos datos se utilizan en el estudio.

Derechos de Autor: Si se utiliza software, bases de datos u otros recursos con derechos de autor en el proceso de análisis de datos, es importante cumplir con las leyes de propiedad intelectual y obtener las licencias necesarias para su uso.

Normativas Comerciales: En el sector retail, existen regulaciones específicas relacionadas con las prácticas comerciales, la publicidad y la competencia leal. Estas regulaciones pueden variar según el país o la región y deben ser consideradas al analizar datos relacionados con ventas y marketing.

Ética de la Investigación: Se deben seguir principios éticos sólidos al llevar a cabo la investigación, incluyendo la obtención de consentimiento informado cuando corresponda, la preservación de la confidencialidad de los datos y la transparencia en la presentación de resultados.

Cumplimiento de Normativas Locales e Internacionales: Dependiendo de la ubicación de la empresa y el alcance de la investigación, puede ser necesario cumplir con normativas locales e internacionales adicionales relacionadas con la gestión de datos y la investigación.

Es fundamental que la investigación se realice de acuerdo con todas las regulaciones y leyes aplicables, y que se obtengan las aprobaciones necesarias si se trata de datos sensibles o de investigación que involucra a seres humanos. Además, se debe documentar adecuadamente el cumplimiento con las normativas legales y éticas en el informe final de la investigación.

CAPÍTULO III. MARCO METODOLÓGICO

Enfoque de la investigación

El enfoque de esta investigación se basa en una perspectiva cuantitativa y descriptiva. Según Hernández, Fernández y Baptista (2014), en el enfoque cuantitativo se busca medir y cuantificar variables para analizar patrones y relaciones en los datos. Dado que se cuenta con una muestra sustancial de 35,000 registros de ventas en el sector retail, se pretende realizar un análisis sistemático y detallado de estos datos.

Por otro lado, el enfoque descriptivo, también propuesto por Hernández, Fernández y Baptista (2014), busca proporcionar una comprensión profunda de los fenómenos estudiados. En este caso, se busca comprender los patrones de compra y los comportamientos de los consumidores en el contexto de ventas en el sector retail. Este enfoque permitirá generar conocimientos sólidos y útiles para la toma de decisiones estratégicas en esta industria, al brindar una visión detallada de la realidad en estudio.

Nivel de la investigación

El presente estudio se sitúa en el nivel descriptivo de la investigación. Este enfoque tiene como objetivo primordial comprender las propiedades, características y perfiles de aspectos específicos en el ámbito de los negocios del sector retail. Tal como señalaron Hernández, Fernández y Baptista (2014), los estudios descriptivos buscan recopilar información sobre conceptos o fenómenos sin la intención de establecer conexiones causales entre ellos. En este contexto, se lleva a cabo un análisis y una medición minuciosa, pero el enfoque no se orienta hacia la identificación de relaciones intrínsecas entre los elementos bajo estudio.

Diseño de la investigación

El diseño de esta investigación se caracteriza por ser no experimental y transversal. Siguiendo las pautas propuestas por Hernández, Fernández y Baptista (2014), se busca obtener información relevante y específica sin la manipulación deliberada de las variables en estudio. En este sentido, no se realizan modificaciones intencionadas a las variables independientes para analizar su efecto sobre otras variables.

La investigación se estructuró como un estudio transversal, lo que significa que se recolectaron los datos en un único momento y no se efectuaron seguimientos a lo largo del tiempo. Este enfoque permite capturar una instantánea representativa de la situación de los datos en el período analizado. De acuerdo con Hernández, Fernández y Baptista (2006), la investigación transversal se enfoca en la recopilación de datos en un punto temporal particular, lo que resulta adecuado para analizar situaciones en un contexto específico sin la necesidad de evaluar cambios a lo largo del tiempo.

Determinación de la población y selección de la muestra

Población

La población de interés en esta investigación está compuesta por los registros de ventas de una empresa del sector retail durante los años 2018, 2019 y 2020. Siguiendo la definición de Hernández, Fernández y Baptista (2010), la población se refiere al conjunto total de casos que cumplen con ciertas especificaciones. En este caso, la población está constituida por todos los registros de ventas de la empresa en los años mencionados.

Es relevante mencionar que la población no solo se limita a un período específico, sino que abarca tres años consecutivos, lo que permite obtener una visión más completa y abarcadora de los patrones de ventas en el sector retail. La selección de estos años específicos se realiza con el propósito de analizar posibles tendencias y variaciones en el comportamiento de las ventas a lo largo del tiempo.

Muestra

Para llevar a cabo esta investigación, se utilizará una muestra compuesta por 35,500 registros de ventas de la empresa del sector retail en los años 2018, 2019 y 2020. La técnica de muestreo empleada será la no probabilística, específicamente la de conveniencia. Esta elección se basa en la disponibilidad y acceso a los datos de ventas de la empresa durante el período establecido.

Conforme a la descripción proporcionada por Hernández, Fernández y Baptista (2014) sobre muestras no probabilísticas, esta técnica permite la selección de elementos basándose en ciertas características relacionadas con la investigación en lugar de depender de la probabilidad. En este caso, la muestra se obtiene de manera conveniente de los registros de ventas disponibles para el período de estudio.

Muestreo

Siguiendo la perspectiva de Hernández Sampieri y su equipo (2010), "una población es el conjunto de todos los casos que concuerdan con una serie de especificaciones".

En este proyecto, se trabajará con una muestra conformada por 35.500 registros de ventas procedentes de una empresa en el sector retail. Estos registros abarcan un período de tiempo específico y serán el foco de análisis para extraer conclusiones relevantes.

Variable de estudio

Hernández, Fernández y Baptista (2014) resaltan que la variable en este contexto se refiere al elemento que se mide, controla y examina dentro del marco del problema planteado. Esta significativa entidad dentro de la investigación posibilita su cuantificación y análisis.

En la presente investigación, las variables de estudio se refieren a los distintos aspectos relacionados con los datos de negocios del sector retail. Las variables abarcarán múltiples aspectos, incluyendo información sobre patrones de compra, características de productos, tendencias de ventas y comportamiento de los consumidores.

Técnica de recolección de datos

Siguiendo las pautas de Sampieri, Collado y Lucio (2013), en esta investigación se empleará una técnica de muestreo no probabilística para seleccionar los 35,500 registros de ventas que formarán parte de la muestra. En este caso, la elección de los elementos de la muestra no se basará en la aleatoriedad, sino en criterios específicos relacionados con las características y objetivos de la investigación. El proceso de selección será intencional y estará guiado por la relevancia de los registros de ventas en el contexto del análisis de datos de negocios en el sector retail.

Para Hernández, Fernández y Baptista (2014) la observación es el método de recolección de datos que consiste en el registro sistemático, válido y confiable de comportamientos y situaciones observables, a través de un conjunto de categorías y subcategorías.

Operacionalización de Variables

Tabla 1. Operacionalización de Variables.

Variable	Definición Operacional	Dimensión	Indicadores	Instrumento
Algoritmos de Clustering	Técnicas de análisis que agrupan datos similares en conjuntos o clústeres	Análisis de Datos de Negocios	Tipos de algoritmos utilizados (p. ej., K-means), eficacia de los algoritmos, capacidad para identificar patrones de compra, velocidad de procesamiento	Revisión bibliográfica, análisis de resultados
Patrones de Compra	Comportamientos recurrentes de los clientes en cuanto a compras	Comportamiento del Consumidor	Frecuencia de compra, productos más comprados, estacionalidad de las compras, canales preferidos para comprar	Ánalysis de datos de ventas, minería de datos
Herramienta KNIME	Plataforma de análisis de datos y minería	Tecnología y Herramientas	Funcionalidades específicas para análisis de negocios, capacidad de procesar grandes volúmenes de datos, facilidad de uso	Revisión de documentación, pruebas de uso

Resultados de Análisis	Conclusiones derivadas de la aplicación de técnicas analíticas	Evaluación de Efectividad	Identificación de segmentos de clientes, recomendaciones de estrategias comerciales, identificación de tendencias	Análisis de resultados, revisión de informes
-------------------------------	--	---------------------------	---	--

Técnicas de procesamiento y análisis de datos

La recolección de datos se llevó a cabo extrayendo información relevante de la base de datos en formato CSV. Posteriormente, estos datos fueron transferidos a una hoja de cálculo en Excel. En la etapa de preparación y análisis, se utilizó la plataforma KNIME para desarrollar un flujo de trabajo específico.

En primera instancia, se configuró el flujo en KNIME para cargar los datos desde el archivo Excel, lo que permitió visualizar y comprender la estructura y contenido de los mismos. Luego, se aplicaron técnicas de limpieza y preprocesamiento de datos, como la identificación y manejo de valores faltantes o atípicos, normalización de variables y eliminación de duplicados. Esto contribuyó a asegurar la calidad y coherencia de los datos.

Una vez que los datos fueron procesados y limpios, se procedió a realizar el análisis utilizando algoritmos de clustering. En este caso, se optó por utilizar el algoritmo K-Means para agrupar los registros en segmentos o clusters relevantes. A través de iteraciones sucesivas, se determinó el número óptimo de clusters, lo que permitió una división adecuada y coherente de los datos.

El resultado del análisis de clustering proporcionó una asignación de registros a diferentes grupos basados en patrones similares. Esto permitió identificar segmentos específicos dentro de los datos de ventas, cada uno caracterizado por patrones de compra similares. Estos hallazgos ofrecen información valiosa para la toma de decisiones estratégicas en el sector retail, al permitir la comprensión de las preferencias y comportamientos de los consumidores.

En resumen, la metodología de recolección de datos involucró la extracción y transformación de información desde una base de datos en formato CSV, seguida por la preparación y análisis de los datos en Excel y KNIME. El uso de técnicas de limpieza,

preprocesamiento y clustering con el algoritmo K-Means contribuyó a la identificación de patrones de compra en los datos de ventas del sector retail. Estos resultados tienen el potencial de informar estrategias y decisiones comerciales informadas y efectivas.

CAPÍTULO IV. ANÁLISIS E INTERPRETACIÓN DE LOS RESULTADOS

Resultados Obtenidos

Dimensión 1. Cantidad de Clusters

Gráfico N° 6. Clusters clasificados por cantidad y visualizado en colores

Cluster
cluster_2
cluster_2
cluster_1
cluster_1
cluster_1
cluster_1
cluster_2
cluster_1
cluster_1
cluster_1
cluster_2
cluster_1
cluster_1
cluster_1
cluster_2
cluster_1
cluster_1
cluster_1
cluster_1

En el análisis de los datos de ventas recopilados durante un período de tres años, se observó una clara tendencia en los productos más vendidos en la empresa. Los resultados revelaron que los productos comestibles ocuparon la posición predominante en términos de ventas, seguidos de cerca por los productos de limpieza.

Productos Comestibles: Durante el período de estudio, los productos comestibles se destacaron como los más populares entre los clientes. Esta categoría incluyó una amplia variedad de alimentos, desde productos enlatados hasta alimentos no perecederos. Los análisis mostraron un aumento constante en las ventas de productos comestibles a lo largo de los tres años, lo que sugiere una fuerte demanda por parte de los clientes.

Productos de Limpieza: Los productos de limpieza ocuparon la segunda posición en términos de ventas. Esta categoría englobó artículos como detergentes, productos para el cuidado del hogar y productos de higiene personal. Aunque las ventas de productos de limpieza fueron significativamente inferiores a las de productos comestibles, se observó una tendencia estable en su demanda a lo largo del período de estudio.

Productos Complementarios: Además de los productos más vendidos, se identificaron ciertos productos complementarios que se destacaron en las transacciones. Estos productos, aunque individualmente no lideraron las ventas, se observó que casi siempre se adquirían junto con otros productos populares. Esta asociación entre productos sugiere la oportunidad de ubicar estos productos complementarios cerca de los productos más vendidos en la tienda, lo que podría aumentar aún más las ventas.

Implicaciones: Estos resultados tienen importantes implicaciones para la estrategia de negocio de la empresa. La alta demanda de productos comestibles sugiere la necesidad de mantener un amplio inventario de estos productos y posiblemente considerar la expansión de esta línea de productos. Al mismo tiempo, la estrategia de ubicar productos complementarios cerca de los productos líderes podría aumentar las ventas y mejorar la experiencia de compra de los clientes.

Relevancia para Futuras Investigaciones: Los resultados obtenidos brindan una visión valiosa de las preferencias de los clientes en relación con los productos vendidos por la empresa. Para futuras investigaciones, sería interesante explorar más a fondo las razones detrás de estas preferencias y cómo pueden influir en la toma de decisiones estratégicas, especialmente en lo que respecta a la disposición de productos en la tienda.

CAPÍTULO V. CONCLUSIONES Y RECOMENDACIONES

Conclusiones

Estado Actual del Sector Retail: La industria minorista es una parte fundamental de la economía, y su funcionamiento eficiente es crucial para mantener un equilibrio en la oferta y la demanda de productos. La investigación reveló que, en el sector retail bajo estudio, existen prácticas tradicionales de gestión de datos que pueden beneficiarse significativamente de enfoques más avanzados y predictivos.

Percepción de la Importancia del Análisis Predictivo: A través de encuestas y análisis de la percepción de los actores clave en la industria minorista, se identificó una creciente conciencia sobre la importancia del análisis predictivo de datos. La mayoría de los encuestados reconocen que esta metodología puede mejorar la toma de decisiones y la eficiencia operativa en el sector.

Comparación de Resultados entre Enfoques Tradicionales y Clustering: La investigación incluyó una comparación entre los métodos tradicionales de análisis de datos utilizados en el sector minorista y el enfoque de clustering. Los resultados demostraron que el clustering permitió una segmentación más precisa de los datos, identificando patrones de compra y comportamientos de los clientes que no eran evidentes mediante métodos convencionales.

Productos Más Vendidos: Uno de los hallazgos más significativos fue la identificación de los productos más vendidos en el sector retail bajo análisis. Estos productos se destacaron por su constante asociación con otros productos específicos, lo que indica la existencia de patrones de compra predecibles. Esta información es valiosa para la planificación de inventario y estrategias de marketing.

Recomendaciones para el Sector: Basado en los resultados del estudio, se formularon recomendaciones para el sector retail. Estas incluyen la adopción de herramientas de análisis predictivo y la implementación de algoritmos de clustering para segmentar a los clientes en grupos con patrones de compra similares. Además, se sugiere que la disposición de ciertos productos cercanos en las tiendas físicas o en las plataformas de comercio electrónico puede aumentar las ventas al aprovechar estos patrones de compra predecibles.

En resumen, esta investigación subraya la importancia del análisis predictivo de datos en el sector retail y demuestra cómo los algoritmos de clustering pueden ser herramientas poderosas para identificar patrones de compra, mejorar la gestión del inventario y aumentar la eficiencia en el negocio minorista. Estos resultados respaldan la idea de que la adopción de enfoques avanzados de análisis de datos puede proporcionar ventajas competitivas significativas en el mercado minorista.

Recomendaciones

A las empresas del sector retail:

Se recomienda utilizar técnicas de análisis de datos avanzadas, como algoritmos de clustering, para identificar patrones de compra entre los productos ofrecidos. Estos patrones pueden ayudar a comprender mejor el comportamiento de los clientes y a tomar decisiones estratégicas, como la disposición de productos en las tiendas o la personalización de ofertas.

A los encargados de la toma de decisiones en la empresa:

Se sugiere que consideren la implementación de un sistema de análisis predictivo basado en algoritmos de clustering para optimizar la gestión de inventario y la planificación de la cadena de suministro. Esto podría conducir a una mejor eficiencia operativa y a una reducción de costos.

A los investigadores y profesionales del análisis de datos:

Se recomienda seguir explorando y desarrollando técnicas de clustering específicas para el sector retail. La aplicación de algoritmos de clustering puede ser de gran utilidad para descubrir información oculta en los datos de ventas y mejorar la toma de decisiones en esta industria en constante cambio.

Referencias

- Agrawal, R., & Srikant, R. (1994). Fast algorithms for mining association rules. Proceedings of the 20th International Conference on Very Large Data Bases, 487-499.
- Han, J., & Kamber, M. (2001). Data mining: concepts and techniques (2nd ed.). Morgan Kaufmann.
- Bhattacharya, S., & Bhattacharya, S. (2014). Customer segmentation and clustering techniques: A review. *International Journal of Business Information Systems*, 17(4), 412-438.
- Jain, S., & Luthra, S. (2016). A comprehensive review on customer purchase behavior analysis in retail sector using data mining techniques. *Procedia Computer Science*, 78, 507-512.
- Wu, D. D., & Olson, D. L. (2010). Business data mining—A machine learning perspective. *Business Intelligence Journal*, 15(1), 15-24.
- Kim, H. W., Lee, H. J., & Lee, I. (2017). Clustering-based customer segmentation approach for personalized recommendation systems. *Journal of Business Research*, 74, 1-10.
- Chen, K., Cheng, K. T., & Chen, T. C. (2013). Customer segmentation using modified

RFM model: A case study on the wireless communication industry. *Expert Systems with Applications*, 40(10), 3976-3985.

Xie, W., & Zhang, Y. (2016). Customer segmentation and target marketing using clustering methods. *Expert Systems with Applications*, 58, 265-273.