

Determination of the Impact on a Supermarket Income Due to the Lack of Products in Gondola by Means of Frequent Consumer Purchases Pattern Recognition

Marcos Moya Navarro, PhD

Universidad Latina de Costa Rica, marcos.moya@ulatina.net

Abstract– The lack of products on the supermarket shelves in Costa Rica had a significant impact on the income of this type of business in 2021. However, only part of the whole products that are displayed on supermarket shelves have an impact on the company's finances as a consequence of consumer preferences. The objective of this work was to determine the economic impact on supermarket finances due to the lack of products in gondola in 2021. The frequent pattern recognition methodology was used as a data mining tool to identify the preferences of clients in the purchase of products and to determine which products are frequently bought together, in such a way that if one of those products is not on the shelf, the customer could leave the supermarket, impacting the income due to the non-sale of all the products in the group. The results of the case study indicated that the projected loss in sales due to the lack of products in gondola is estimated to be close to forty-six thousand three hundred and seventy-four dollars per year, with a projected decrease in supermarket profits of twelve thousand six hundred and eighty-four dollars.

Keywords-- Supply Chain Management, Data Mining, Pattern Recognition, Business Intelligent.

Digital Object Identifier (DOI):

<http://dx.doi.org/10.18687/LACCEI2022.1.1.515>

ISBN: 978-628-95207-0-5 **ISSN:** 2414-6390

Determinación del Impacto en los Ingresos de un Supermercado por Falta de Productos en Góndola Mediante el Reconocimiento de Patrones de Compras Frecuentes del Consumidor

Determination of the Impact on a Supermarket Income Due to the Lack of Products in Gondola by Means of Frequent Consumer Purchases Pattern Recognition

Marcos Moya Navarro, PhD
Universidad Latina de Costa Rica, marcos.moya@ulatina.net

Resumen– La falta de productos en los anaqueles de los supermercados en Costa Rica tuvo un impacto significativo en los ingresos de este tipo de negocios en el 2021. Sin embargo, solo una parte del total de los productos que se exhiben en los anaqueles de los supermercados tienen un impacto en las finanzas de la empresa como consecuencia de las preferencias de los consumidores. El presente trabajo tuvo como objetivo determinar el impacto económico en las finanzas de un supermercado por la falta de productos en góndola en el año 2021. Se utilizó la metodología de reconocimiento de patrones frecuentes como herramienta de minería de datos para identificar las preferencias de los clientes en la compra de productos y determinar qué productos se compran juntos con frecuencia, de tal manera que, si uno de esos productos no está en el anaquel, el cliente podría salir del supermercado, impactando los ingresos por la no venta de todos los productos del grupo. Los resultados del caso de estudio indicaron que la pérdida proyectada en ventas por falta de productos en góndola se estima cercana a los cuarenta y seis mil trescientos setenta y cuatro dólares anuales, con una disminución proyectada en las utilidades de los supermercados de doce mil seiscientos ochenta y cuatro dólares.

Palabras Clave-- Gestión de la Cadena de Suministro, Minería de Datos, Reconocimiento de Patrones, Inteligencia de Negocio.

Abstract– The lack of products on the supermarket shelves in Costa Rica had a significant impact on the income of this type of business in 2021. However, only part of the whole products that are displayed on supermarket shelves have an impact on the company's finances as a consequence of consumer preferences. The objective of this work was to determine the economic impact on supermarket finances due to the lack of products in gondola in 2021. The frequent pattern recognition methodology was used as a data mining tool to identify the preferences of clients in the purchase of products and to determine which products are frequently bought together, in such a way that if one of those products is not on the shelf, the customer could leave the supermarket, impacting the income due to the non-sale of all the products in the group. The results of the case study indicated that the projected loss in sales due to the lack of products in gondola is estimated to be close to forty-six thousand three hundred and seventy-four dollars per year,

with a projected decrease in supermarket profits of twelve thousand six hundred and eighty-four dollars.

Keywords-- Supply Chain Management, Data Mining, Pattern Recognition, Business Intelligent.

I. INTRODUCCION

Con la enfermedad del COVID 19 se modificaron los hábitos del consumidor en Costa Rica durante el año 2021 de manera significativa. La Ref. [1] indica que la fidelidad de los consumidores de Costa Rica a una marca o cadena de supermercados depende de los precios de los productos que ofrezca, las promociones que realice y la ubicación o cercanía del punto de venta a los hogares, al tiempo que se incrementa la compra en línea especialmente entre los compradores más jóvenes. El autor también indica que un 63% de los consumidores están dispuestos a cambiar sus hábitos de visitación y sus hábitos de compra y consumo a partir de lo que encuentren en los sitios de comercio electrónico de las cadenas de supermercados.

La Ref. [2] establece que el hábito de fraccionar las compras en 2 supermercados presenta una caída importante de 10 puntos porcentuales, pasando de un 49% en el 2020 a un 39% de los consumidores en el año 2021. El autor indica además que esta disminución permite que ahora el 43% de los consumidores manifiesten que concentran su compra en un solo comercio ya que este les brinda todo lo que andan buscando con respecto a mejores precios, promociones y surtido/variedad de productos. Esta es una de las razones que justifican la importancia de la disponibilidad de productos ya sea en los anaqueles de los supermercados o bien en las bodegas de las tiendas.

La Ref. [3] indica al respecto de la disponibilidad de productos en góndola y hábitos de compra del consumidor que, si el producto no está disponible, los clientes irán a otro lugar o comprarán un producto de los competidores. Este factor impacta directamente las finanzas de los supermercados

Digital Object Identifier (DOI):

<http://dx.doi.org/10.18687/LACCEI2022.1.1.515>

ISBN: 978-628-95207-0-5 ISSN: 2414-6390

en sus ingresos por ventas y en sus márgenes de utilidad. Estudios preliminares realizados por GS1 Costa Rica indican que, sobre la base de 109 000 mediciones, el faltante proyectado de mercadería en góndola en supermercados ascendió al cierre de diciembre del 2021 a un 5.68%. Este indicador obtuvo una mejoría significativa con respecto al 2020, año en el cual el porcentaje de faltantes fue de 9.16%, Ref. [4], lo cual implica una mejora del indicador del 38%.

La Fig.1 muestra el comparativo anual que indica que durante el año 2021 el 41% de los consumidores tomaron la decisión de adquirir el producto que buscaba en otro lugar, cuando no encuentra disponible en el supermercado el producto de la marca que desean adquirir. La figura también indica que este porcentaje de consumidores se ha mantenido prácticamente estable desde el año 2018.

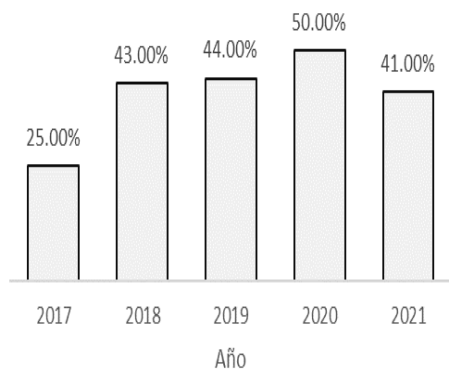


Fig. 1. Decisión de adquirir el producto de la marca que buscaba en otro lugar cuando esa marca no está disponible en el supermercado
Fuente: GS1 Costa Rica.

La Ref. [5] indica que no tener los productos en góndola es un error que cuesta 31.000 millones de dólares a las compañías de comercio mayorista, según un estudio de la firma Logyca. El estudio incluyó el 80% de las ventas que más rotan y 60% de los puntos de venta que tienen bajo su registro. Este estudio demuestra el impacto económico de los problemas de logística que enfrentan las compañías de comercio mayorista.

La Ref. [6] establece que más de dos tercios de los compradores abandonaron una tienda porque no tenían el producto que querían. El autor indica que el resultado es pérdida de ventas, existencias potencialmente desperdiciadas y una mella en el resultado final. El autor indica que la propagación de COVID-19 en todo el mundo ha puesto de relieve la disponibilidad en los estantes, ya que los minoristas de comestibles luchan por mantenerse al día.

La Ref. [7] indica que los consumidores siguen prefiriendo comprar en la tienda, pero la baja disponibilidad

de productos en góndola y la gran cantidad de productos agotados pueden afectar la experiencia del cliente minorista. El autor establece que, si los consumidores no encuentran lo que buscan en los estantes, el 49 por ciento elegirá un producto sustituto, el 39 por ciento se irá con las manos vacías y el 12 por ciento se dirigirá a un competidor para satisfacer sus necesidades, por lo que en cada caso el minorista, el fabricante y el cliente pierden.

La Ref. [8] indica que los minoristas estadounidenses perdieron \$82 mil millones en ventas de bienes de consumo envasados porque los artículos que los clientes habrían comprado no estaban disponibles, según datos publicados el lunes por NielsenIQ. El autor también establece que las oportunidades de ingresos perdidos alcanzaron su punto máximo durante el otoño, cuando las crecientes tasas de COVID-19 redujeron la disponibilidad en el estante y llevaron las ventas no realizadas a \$ 1.7 billones de dólares por semana, un 22% más que los 1.4 billones de dólares por semana durante la primera mitad de 2021. Aun más, el autor señala que las cifras subrayan el impacto que los desafíos de la cadena de suministro han tenido en los minoristas de alimentos, incluso cuando la pandemia ha hecho que los consumidores dependan más de las tiendas de comestibles para obtener productos esenciales.

La Ref. [9] establece que, en enero de 2016, la escasez de productos en las estanterías de los grandes establecimientos italianos provocó una pérdida de ventas del 4,2% del volumen total de ventas, por lo que estas cifras plantean una serie de preguntas, tanto sobre el impacto financiero de la falta de existencias para los minoristas y fabricantes, así como sobre cómo esto afecta el comportamiento de los consumidores, ya que no encuentran lo que necesitan en los estantes. El autor además indica que en el año 2015 la pérdida de ventas provocada por productos no disponibles en las góndolas se estimó en unos 700 millones de euros y que las ventas incompletas se debieron a que los consumidores se abstienen de comprar ya que los minoristas no pueden satisfacer sus necesidades. Continúa indicando el autor que el 2015 se agotaron el 3,7% de los productos, 4% en el 2014 y se perdieron el 4,7% de las ventas frente al 5% de desabastecimiento alcanzado el año anterior.

La Ref. [10] indica que en el año 2003, se encargó al profesor Daniel Corsten y Thomas Gruen que reunieran los datos de más de 50 estudios para producir una visión global del impacto de la falta de existencias en las ventas y las ganancias. Los autores determinaron que el minorista incurriría en una pérdida de ventas cuando el comprador realiza una de las dos acciones siguientes: No compra nada o bien compra el artículo en otra tienda. También estimaron, evidenciando un estudio de Data Ventures, una pérdida de valor del 7 por ciento cuando el comprador sustituye artículos y compra marcas alternativas a un precio minorista más bajo o tamaños más pequeños de la misma marca. En consecuencia,

juntas, estas pérdidas reales suman un 43 por ciento, lo que significa que, si los compradores se enfrentaran a 100 incidentes de falta de existencias en un estante, 43 de esos incidentes generarían menos dinero en la caja, y el minorista no perdería valor de ventas inmediato en los otros 57 incidentes cuando los consumidores cambiaron a otras marcas.

A. Inteligencia de Negocios

El objetivo de este trabajo consiste en determinar el impacto económico en las finanzas de un supermercado por la falta de productos en góndola en el año 2021. Debido a la pandemia del COVID 19 la digitalización de los negocios sufrió un cambio radical y significativo en Costa Rica, lo cual implicó que el volumen de datos generados creciera exponencialmente. Por esta razón, el uso de herramientas de inteligencia de negocio se vuelve útiles para conocer las preferencias de los consumidores a partir de los grandes volúmenes de datos de las transacciones que estos clientes generan, y poder así determinar el impacto económico en que incurren los negocios cuando no se dispone en los estantes del comercio los productos que buscan esos clientes.

La Ref. [11] indica que el ambiente en el que actualmente las compañías se desenvuelven es de alta competencia y requiere que los gerentes tomen decisiones rápidas y concisas, obligándolos a buscar tanto soluciones como estrategias que generen la tan preciada ventaja competitiva. El autor también indica que la tecnología al estar en constante crecimiento juega un rol preponderante para la toma de decisiones, y que los avances en el área permiten a las empresas tener las herramientas necesarias para hacer frente a sus problemas de negocios. En esta coyuntura, los sistemas de Business Intelligence están diseñados para apoyar a las personas encargadas de tomar decisiones en las empresas.

La Ref. [12] indica que la inteligencia comercial (BI) ahora se usa ampliamente, especialmente en el mundo de la práctica, para describir aplicaciones analíticas y es actualmente la principal prioridad de muchos directores de información. El autor indica que la inteligencia de negocios se ha convertido en una iniciativa estratégica y ahora es reconocida por los directores y los líderes empresariales como fundamental para impulsar la eficacia y la innovación empresarial.

La Ref. [13] indica que se puede evaluar el comportamiento y gestión de los puntos de venta de una compañía y de su competencia a través de la minería de datos, la cual es una herramienta de inteligencia de negocios. El autor establece que la minería de datos es una herramienta que permite explorar información útil de bases de datos de manera automática o semi automática, con el fin de encontrar patrones repetitivos o tendencias ocultas que expliquen el comportamiento de un grupo de datos en particular. El autor también destaca que la materia prima para el desarrollo de la

minería de datos son las bases de datos, las cuales deben ser confiables. Para determinar el impacto económico de los productos faltantes en góndola, en este estudio se utilizó una base de datos de 20 000 registros de los tiquetes de compra de clientes durante el año 2021. El estudio de campo fue realizado por la organización sin fines de lucro GS1 Costa Rica.

La Ref. [14] establece que la inteligencia de negocios (BI) se refiere a las aplicaciones y tecnologías que se utilizan para recopilar, proporcionar acceso o analizar información sobre las operaciones de una empresa. El autor indica que los sistemas de Business Intelligence ayudan a las empresas a tener un conocimiento integral de los factores que afectan su negocio; por ejemplo, métricas de ventas, producción y operaciones internas. El autor también explica que las aplicaciones de BI incluyen herramientas de informes y análisis de consultas, herramientas de minería de datos y herramientas de almacenamiento de datos.

La Ref. [15] indica que la minería de datos se puede utilizar como una herramienta de inteligencia empresarial para el descubrimiento de conocimientos (Wang y Wang, 2008) y se emplea cada vez más en una variedad de dominios ricos en datos (Weiss et al., 2008). El autor establece que los algoritmos basados en minería de datos a menudo se han aplicado a empresas de servicios comerciales para descubrir patrones desconocidos e información oculta de grandes bases de datos. Por lo tanto, descubrir un conjunto de elementos “interesantes” de una gran base de datos de transacciones de empresas se ha convertido en una de las principales tareas para los tomadores de decisiones en la investigación reciente de minería de datos (Lee et al., 2013).

B. Metodología Para el Reconocimiento de Patrones de Preferencias de Compra de los Consumidores

Se utilizó la metodología de reconocimiento de patrones frecuentes como una herramienta de la minería de datos para identificar las preferencias de los clientes en la compra de productos en un supermercado y determinar así qué productos se compran juntos frecuentemente, de tal manera que, si uno de esos productos no está en el anaquel, el cliente podría salir del supermercado, impactando los ingresos por la no venta de todos los productos del grupo.

La Ref. [16] establece que existen tres metodologías básicas para el reconocimiento de patrones: las heurísticas, las matemáticas y las lingüísticas. El autor indica que en las metodologías heurísticas se hace uso de la experiencia y la intuición humana. En las metodologías matemáticas se hace uso de las propiedades comunes de los patrones y se basan en las reglas de clasificación formuladas en un marco matemático. En las metodologías lingüísticas se hace uso de

los elementos primitivos que componen a los patrones llamados subpatrones y la relación que existe entre ellos.

La Fig. 2 resume las etapas que es necesario llevar a cabo para resolver problemas de reconocimiento de patrones.

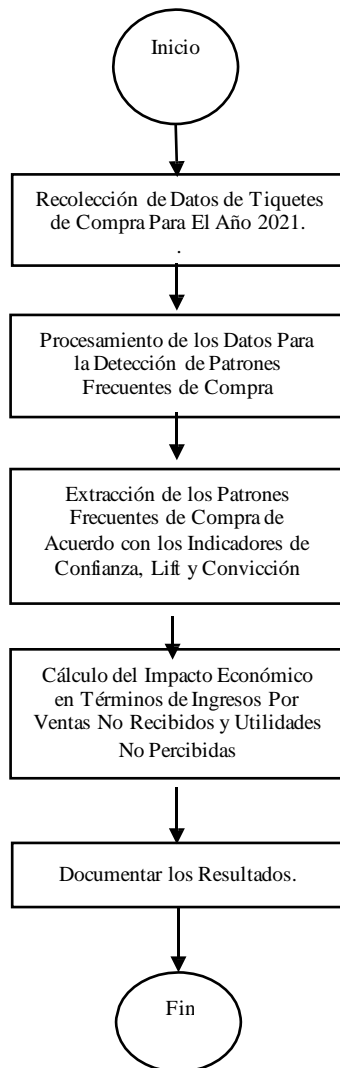


Fig. 2. Etapas para el reconocimiento de patrones.

La Ref. [17] indica que la disciplina denominada Minería de Datos estudia métodos y algoritmos que permiten la extracción automática de información sintetizada que permite caracterizar las relaciones escondidas en la gran cantidad de datos, pretendiendo también que la información obtenida posea capacidad predictiva. El autor también apunta que este tipo de herramientas de descubrimiento de conocimiento contenido en las bases de datos establece las posibles relaciones o correlaciones entre distintas acciones o sucesos aparentemente independientes, pudiendo reconocer como la ocurrencia de un suceso o acción puede inducir o generar la aparición de otros.

Para la solución del problema se utilizó una base de datos de las compras que hicieron los clientes en un supermercado durante el año 2021. Para efectos de determinar los patrones de compra la base de datos contiene los tiquetes de compra que indican si un determinado producto se compró. Si hubo compra de ese producto el campo muestra un 1. Si no hubo compra de ese producto, la base de datos muestra un cero.

La Ref. [18] indica que los patrones de compra se refieren a las decisiones de compra de los consumidores. Son hábitos y rutinas que los consumidores establecen a través de los productos y servicios que compran. El autor establece que los patrones de compra se definen por la frecuencia, el tiempo, la cantidad, etc.... de dichas compras, y que estos patrones están determinados por factores como lugar de residencia, lugar de trabajo, Ocio y hobbies, donde trabajan, lo que recomiendan sus amigos y familiares, cuáles son sus objetivos y motivaciones, el precio del producto o servicio que les interesa, la necesidad de un producto o servicio, y festivales, días festivos o celebraciones.

Con respecto a la extracción de los patrones frecuentes de compra se utilizó el algoritmo aPriori (Frequent Itemset Mining Algorithm), el cual es apropiado para determinar un conjunto de itemsets frecuentes en la base de datos.

La Ref. [19] indica que la minería de patrones frecuentes es una de las técnicas más importantes de minería de datos para descubrir relaciones entre diferentes elementos en un conjunto de datos. Estas relaciones se representan en forma de reglas de asociación. El autor establece que se requieren dos conjuntos fundamentales para determinar las relaciones entre los diferentes elementos en la base de datos: Un conjunto de “n” atributos binarios llamados elementos y un conjunto de transacciones llamado base de datos. Una regla de asociación se define como una implicación de la forma X implica Y o bien Y implica X.

Para determinar los patrones frecuentes de compra se requiere especificar un umbral de soporte. Este es el número mínimo de transacciones que contienen el itemset x dividido por el total de transacciones. La Ref. [20] especifica que el indicador de “Confianza” se determina dividiendo el soporte de la unión(XY), entre el soporte de X. La unión(XY) es el itemset que contienen todos los ítems de X e Y. Este indicador se interpreta como la probabilidad que se tiene de comprar un ítemset Y dado que se compró el ítemset X o viceversa.

La Ref. [21] establece que el indicador “Lift” se interpreta como el aumento de la probabilidad de venta del itemset “Y” si se sabe que el cliente compro el itemset “X”. El Lift también puede interpretarse como el grado en que el itemset Y tiende a ser frecuente cuando el itemset X Ocurre.

La Ref. [21] también establece que el indicador de “Convicción” es una medida que tiene el objetivo de evaluar el grado en que el antecedente influye en la ocurrencia de consecuente. Esto es el grado en que la compra del itemset “X” influye en la compra del itemset Y.

II. DEFINICION DEL CASO DE ESTUDIO

Un estudio realizado por la organización sin fines de lucro GS1 Costa Rica buscó medir los hábitos del consumidor en el año 2021. Participaron en ese estudio cinco cadenas del canal moderno de supermercados en los formatos de hipermercado, supermercado, bodegas y supermercados de descuento. Se realizaron 15 preguntas a los consumidores a la salida de los supermercados. La Fig.3 resume las preguntas realizadas.

Preguntas realizadas a los consumidores:

1. ¿ Por qué razón escoge este punto de venta?
2. ¿Cuál es su frecuencia de compra del diario?
3. ¿Habitualmente donde realiza usted la compra del diario?
4. ¿Por qué motivo fracciona usted su compra?
5. ¿Qué hace usted cuando no encuentra disponibles los productos de las marcas que deseaba adquirir?
6. ¿Compró algún producto que no tenía programado comprar?
7. ¿Qué influyo para hacer la compra de lo que no tenía planeado?
8. ¿Las cosas que se le van agotando de su diario, donde las compra?
9. ¿Visita alguna tienda de conveniencia?
10. ¿Cuál es su principal método de pago en el supermercado?
11. ¿ Dónde compra las frutas y verduras?
12. ¿ Dónde compra las carnes?
13. ¿Realiza usted compras en línea?
14. ¿Según su frecuencia, qué tipo de productos compra en línea?
15. ¿En el futuro estaría usted dispuesto a comprar su diario en línea?

Fig. 3. Preguntas realizadas a los consumidores a la salida de los supermercados.

Fuente: GS1 Costa Rica.

Una pregunta que destaca es la pregunta 5: ¿Qué hace usted cuando no encuentra disponibles los productos de las marcas que deseaba adquirir?

El estudio indicó que el 26% de los clientes no realiza la compra del producto, el 33% adquiere el mismo producto, pero de otra marca, y el 41% adquiere el producto que buscaba en otro lugar. Es importante destacar que el 15.42% de los consumidores indicaron no haber encontrado todos los productos que buscaban. Estos resultados evidencian el impacto significativo para el supermercado en términos económicos por los ingresos por ventas que dejan de percibir y que impactan de manera directa las utilidades del negocio.

Para este estudio se utilizó una base de datos de 20000 registros de tiquetes de compra durante el año 2021 de las compras registradas en los tiquetes de caja de un supermercado situado en la provincia de Guanacaste, Costa Rica. La base de datos incluye 6989 productos, en las

categorías de abarrotes y cuidado personal, mercadería general y productos frescos (carnes y embutidos, frutas y verduras, y panadería). La Fig.4 resume la composición de productos de la base de datos por categorías de productos.

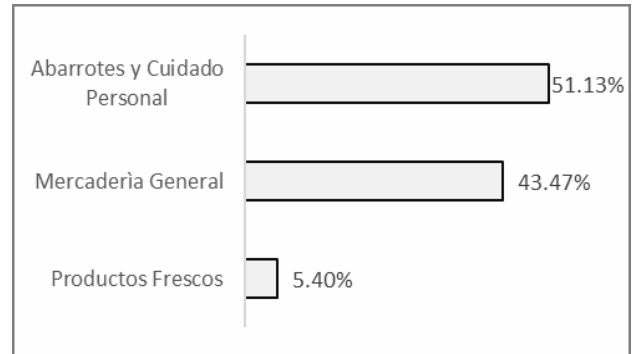


Fig. 4. Categorías de productos de la base de datos en estudio para determinar los Itemsets frecuentes de compras.

Fuente: GS1 Costa Rica.

El objetivo de este trabajo consistió en determinar el impacto económico en las finanzas de un supermercado por la falta de productos en góndola en el año 2021. Para ello, se utilizó la metodología de reconocimiento de patrones frecuentes como herramienta de minería de datos para identificar las preferencias de los clientes en la compra de productos y determinar qué productos se compran juntos con frecuencia, de tal manera que, si uno de esos productos no está en el anaquel, el cliente podría salir del supermercado, impactando los ingresos por la no venta de todos los productos del grupo.

III. METODOLOGIA DE RECONOCIMIENTO DE PATRONES UTILIZANDO RAPIDMINER

La etapa de extracción de itemsets frecuentes se llevó a cabo mediante la utilización del programa informático RapiMiner Studio (2001-2021), el cual es una marca registrada de RapidMiner GbmH.

La Fig. 5 resume la metodología de extracciones de ítems frecuentes y la creación de las reglas de asociación mediante el uso del programa informático RapidMiner Studio.

El proceso de extracción de ítems frecuentes realizado por el programa informático RapidMiner incluye cuatro subprocesos. En el subproceso 1 se carga la base de datos. En el subproceso 2, se convierte el número de productos comprados, según se indica en el tiquete de compra correspondiente, en variables binarias verdadero y falso, que indican si el producto se compró o no. En el subproceso 3 se extraen los ítems frecuentes, utilizando los algoritmos

apropiados de extracción. Se requiere especificar un umbral mínimo de soporte, el cual está asociado a la cantidad de ítems adquiridos por compra. Finalmente, en el subproceso 4 se crean las reglas de asociación para los ítems frecuentes identificados en el subproceso 3. Se requiere especificar una confianza mínima requerida.

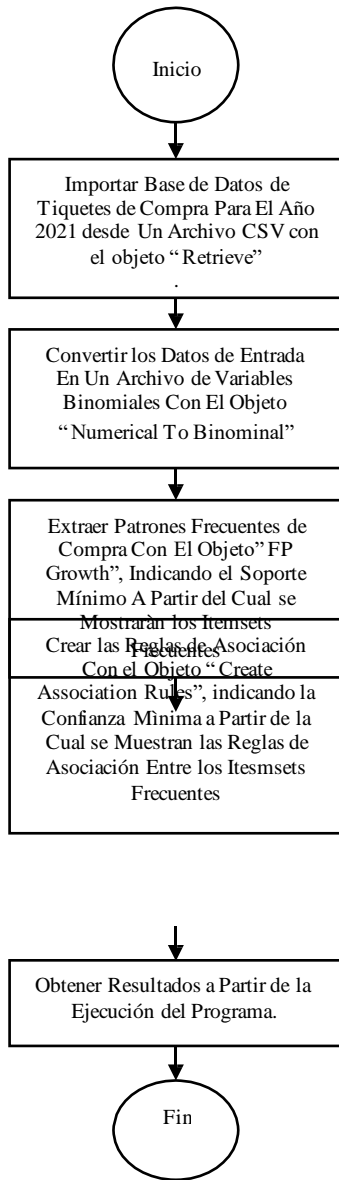


Fig. 5. Procedimiento para el reconocimiento de patrones con el Programa informático RapidMiner Studio.

IV. IDENTIFICACION DE LOS ITEMSETS FRECUENTES Y CALCULO DEL IMPACTO ECONOMICO

La Fig. 6 muestra el proceso completo para identificar los ítems de productos frecuentes en la base de datos. Los bloques clave para este proceso son “Retrieve”, “Numerical to Binominal”, “FP Growth” y “Create Association Rules”

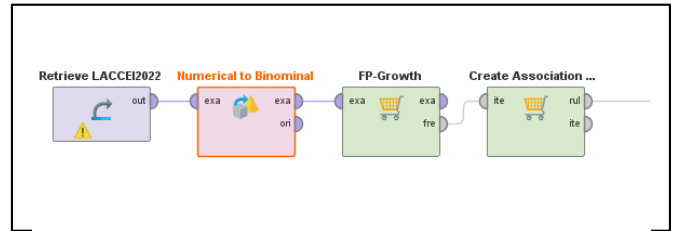


Fig. 6. Bloques de construcción del árbol de decisión para la determinación de ítems frecuentes con RapiMiner Studio.

La Tabla I muestra un segmento de la base de datos de compras de los clientes, importada mediante el bloque de importación “Retrieve”

TABLA I
SEGMENTO DE LA BASE DE DATOS DE COMPRAS DE CLIENTES

UltDia_id	tiquete_id	HIELO CRISTAL 4KG	HIELO CRISTAL EN MARQUETA
1/1/2021	33278820115	1	0
1/1/2021	33284390115	1	0
1/1/2021	33327270115	0	1
1/1/2021	33327430115	0	1
1/1/2021	33327580115	0	0
1/1/2021	33327750115	0	0
1/1/2021	33327800115	0	0
1/1/2021	33327980115	0	0
1/1/2021	33328140115	0	0
1/1/2021	33328190115	0	0
1/1/2021	33328260115	0	0
1/1/2021	33328300115	0	0
1/1/2021	33328420115	0	0
1/1/2021	33328560115	0	0
1/1/2021	33328650115	1	0

La Tabla II muestra un segmento de los resultados obtenidos para los ítems frecuentes que se identificaron a partir de una confianza especificada de 90%. El antecedente representa todos aquellos productos que se compraron juntos y que influyen de manera contundente en la compra del consecuente. Como ejemplo para el primer ítemset frecuente se tiene que la venta del producto Sandalia para Niña Carioca está asociada significativamente por la venta de los productos Salsa Picante Jalapeño 150 gramos, vino Contenda Sauvignon Blanco, y sandalia Mujer Brasil #39. En consecuencia, si los productos del antecedente no están disponibles en la góndola, el cliente puede abandonar el supermercado afectando la venta de los productos que están asociados en el consecuente.

El mismo tipo de análisis debe de extenderse se para todos y cada uno de los consecuentes encontrados en el análisis.

TABLA II
SEGMENTO DE RESULTADOS DE LOS ITEMSETS
FRECUENTES OBTENIDOS PARA UNA
CONFIANZA DE 90%

Antecedente	Consecuente
SALSA PICANTE JALAPEÑO 150, COTENDA SAUVIGNON BLANC, SANDALIA MUJER BRAZIL #39, SALSA SWEET HONEY, COTENDA SAUVIGNON BLANC, COTUBA CERO 2L	SANDALIA P/NIÑA CARIOCA
SALSA PICANTE JALAPEÑO 150, COTUBA CERO 2L, CORREA RETRACTIL 8MTS,	SANDALIA P/NIÑA CARIOCA
SANDALIA P/NIÑA CARIOCA, SALSA TOMATE HONGOS, COTUBA CERO 2L, CORREA RETRACTIL 8MTS	SANDALIA P/NIÑA CARIOCA COTENDA SAUVIGNON BLANC
SALSA PICANTE JALAPEÑO 150, DOS PINOS JUGO NARANJA TETRA 250ML, CORREA RETRACTIL 8MTS,	SANDALIA P/NIÑA CARIOCA
SALSA PICANTE JALAPEÑO 150, SALSA TOMATE HONGOS, COTENDA SAUVIGNON BLANC, COTUBA CERO 2L	SANDALIA P/NIÑA CARIOCA
...	...
SANDALIA P/NIÑA CARIOCA, SALSA SWEET HONEY, COTENDA SAUVIGNON BLANC, COTUBA CERO 2L	SALSA TOMATE HONGOS
SALSA TOMATE HONGOS, CORONADO LIQUIDO FRUTAS TROPICALES BOTELLA 1000ML, SALSA PICANTE MEAN GREEN, COTUBA CERO 2L	CORREA RETRACTIL 8MTS
SALSA KETCHUP PLASTICO 567ML, CORONADO LIQUIDO FRUTAS TROPICALES BOTELLA 1000ML, SALSA PICANTE MEAN GREEN, CORREA RETRACTIL 8MTS	SALSA TOMATE HONGOS
SALSA KETCHUP PLASTICO 567ML, SALSA PICANTE MEAN GREEN, COTUBA CERO 2L, CORREA RETRACTIL 8MTS	SALSA TOMATE HONGOS
SALSA KETCHUP PLASTICO 567ML, CORREA RETRACTIL 8MTS, CIGAR PALL MALL IBIZA FRES,	SALSA TOMATE HONGOS

TABLA III
SEGMENTO DE RESULTADOS DE LOS INDICADORES
DE DESEMPEÑO DE LAS REGLAS DE ASOCIACION
DE LOS ITEMSETS FRECUENTES

Consecuente	Confianza	Lift	Convicción
SANDALIA P/NIÑA CARIOCA	0.911	12.719	10.398
SANDALIA P/NIÑA CARIOCA	0.912	12.740	10.584
SANDALIA P/NIÑA CARIOCA	0.913	12.751	10.677
COTENDA SAUVIGNON BLANC	0.913	39.787	11.236
SANDALIA P/NIÑA CARIOCA	0.915	12.777	10.909
SANDALIA P/NIÑA CARIOCA	0.918	12.821	11.326
SALSA KETCHUP PLASTICO 567ML	0.918	20.288	11.695
SALSA KETCHUP PLASTICO 567ML	0.918	20.288	11.695
...
SALSA TOMATE HONGOS	0.978	21.170	42.922
SALSA TOMATE HONGOS	0.978	21.170	42.922
SALSA TOMATE HONGOS	1.000	21.651	Infinity

La Tabla III muestra los indicadores de desempeño de las reglas de asociación obtenidas por el programa de minería de datos RapidMiner. Los itemsets seleccionados fueron aquellos que tuvieran al menos una confianza del 90%, siendo la menor confianza obtenida de 91.11%. Esto significa que se estima una probabilidad del 91.11% de comprar la sandalia para niña carioca dado que se compraron los productos salsa picante jalapeño 150 gramos, vino contenda Sauvignon blanco y sandalia de mujer brazil #39. Aún más, se puede indicar que el 91.11% de los clientes que compraron salsa picante jalapeño 150 gramos, vino contenda Sauvignon blanco y sandalia de mujer brazil #39 también compraron sandalia para niña carioca.

El indicador Lift se interpreta como el aumento de probabilidad del consecuente si se sabe que se compró el antecedente. Esto significa que para el primer itemset frecuente de la Fig. 8 la probabilidad de comprar sandalias para niña carioca se aumenta alrededor de doce veces si se sabe que el cliente compró salsa picante, vino Sauvignon blanco y sandalia de mujer brazil.

Finalmente, el indicador de “Convicción” muestra el grado en que la compra del itemset “X” influye en la compra del itemset Y. Para el primer itemset frecuente de la Fig.8, se muestra en la Fig. 9 que la sandalia para niña carioca está siendo influenciada de manera significativa por la compra de la salsa picante jalapeño 150 gramos, el vino Sauvignon blanco y la sandalia de mujer brazil #39. Como se indicó antes, se debe extender el mismo razonamiento para los demás itemsets frecuentes en la tabla de resultados. Entre mas grande sea el indicador de convicción, mayor es la influencia que tienen los productos del antecedente en los productos del consecuente.

V. IMPACTO ECONOMICO DE LOS ITEMSETS FRECUENTES

El estudio indicó que para el primer itemset frecuente de la Tabla II y su correspondiente consecuente producto sandalia para niña carioca se proyecta un ingreso anual no percibido por el supermercado de un mil nueve dólares americanos si los clientes no lo encuentran en góndola y buscan los productos en la competencia. Este hecho impactaría la utilidad del negocio en trescientos noventa dólares americanos por la no venta de estos productos. Para la totalidad de los itemsets frecuentes determinados, se proyecta un ingreso no percibido de cuatrocientos sesenta y tres mil setecientos cuarenta dólares, lo que impactaría la utilidad del supermercado en ciento veintiséis mil ochocientos cuarenta dólares americanos.

CONCLUSIONES

Este trabajo utilizó el reconocimiento de patrones como método de minería de datos para determinar la influencia que tienen un grupo de productos que se compran con frecuencia en la compra de otros productos que también se compran con frecuencia.

Las principales conclusiones obtenidas son las siguientes.

1. Los itemsets frecuentes están determinando por el comportamiento de los consumidores y pueden variar de un establecimiento a otro, según zona geográfica, formato del supermercado, estacionalidades de los productos, y la variedad de productos.
2. El impacto económico para los establecimientos depende de la composición de los itemsets frecuentes determinados.
3. Los algoritmos de minería de datos no supervisados son herramientas muy apropiadas para el reconocimiento de patrones para la determinación de los itemsets frecuentes.
4. Para el supermercado en estudio se proyectó un ingreso anual no percibido de 463740 dólares americanos, impactando la utilidad del negocio en 126840 dólares.

Finalmente, una línea de acción futura será extender este tipo de estudios para distintos formatos de supermercados.

AGRADECIMIENTO

Un reconocimiento y agradecimiento a la organización sin fines de lucro GS1 Costa Rica, y en especial al Ing. José Miguel Ugalde Rodríguez quien proporcionó y extrajo los datos apropiados de las bases de datos para que este trabajo se pudiera llevar a cabo. También un agradecimiento a la empresa RapidMiner quien proporcionó una licencia académica del programa computacional de minería de datos RapiMiner Studio. Sin esta herramienta no hubiera sido posible la realización de este trabajo.

REFERENCIAS

- [1] Cordero, C. Precios. Promociones y cercanía determinan fidelidad de los clientes de supermercados, Noviembre, 2021. <https://www.nacion.com/economia/consumo/precios-promociones-y-cercania-determinan/EVAYSK7DKRBUP07HOC2JOJXL7A/story/>
- [2] Revista Summa. ¿Cuáles son los hábitos de consumo de los costarricenses? <https://revistasumma.com/cuales-son-los-habitos-de-consumo-de-los-costarricenses/>
- [3] GS1 Latam. Estudio de disponibilidad de mercadería en góndola y hábitos de compra del consumidor. 2019. <http://docplayer.es/179946102-Xx-estudio-de-disponibilidad-de-mercaderia-en-gondola-y-habitos-de-compra-del-consumidor.html>
- [4] Moya, M. Identificación De Los Principales Factores Que Influyen En El Faltante De Productos En Las Góndolas De Los Supermercados Por Medio Del Análisis Factorial Por Componentes Principales.

- <https://laccei.org/LACCEI2021-VirtualEdition/meta/FP181.html>
- [5] Grupo Semana. No tener los productos en la góndola: un error que cuesta \$31.000 millones. 2022. <https://www.semana.com/economia/articulo/impacto-de-no-tener-los-productos-en-gondola-para-los-supermercados/273288/>
- [6] Retail Insight. Why do grocery retailers find on-shelf availability such a challenge? 2022. <https://www.retailinsight.io/blog/why-do-grocery-retailers-find-on-shelf-availability-such-a-challenge>
- [7] Trax Retail. Why on-shelf availability is still an issue (and how to address it). 2019. <https://traxretail.com/blog/shelf-availability-still-issue-address/>
- [8] Silverstein, S. Empty shelves sapped retailers of \$82B in CPG sales last year. 2022. <https://www.grocerydive.com/news/empty-shelves-sapped-retailers-of-82b-in-cpg-sales-last-year/618569/>
- [9] thena Retail. Empty shelves? The result: lost sales and dissatisfied customers. 2016. <https://www.athenaretail.com/en/empty-shelves-the-result-lost-sales-and-dissatisfied-customers/>
- [10] Peacot, C. Fixing Shelf Out-of-Stocks. Loss Prevention Magazine Europe. 2022. <http://www.thelpportal.com/content/editorial/articles/retail-environment/fixing-shelf-out-stocks/>
- [11] Adasme, D, Salomón, E. Análisis de Patrones de Compras de Tiendas Retail utilizando Business Intelligence. Facultad de Economía y Negocios. Universidad de Chile. 2013. <https://repositorio.uchile.cl/bitstream/handle/2250/115078/Adasme%20A.%2C%20Daniel.pdf?sequence=1&isAllowed=y>
- [12] Watson, Hugh J; Wixom, Barbara H. The Current State of Business Intelligence. Computer; New York Tomo 40, N.º 9, 96-99. 2007
- [13] Centro de Investigación de Mercados. Minería de Datos. <https://www.ciminvestigacion.com/datamining-mineria-de-datos/>
- [14] CMI Close Up Media. Research and Markets Adds Report: Global Business Intelligence Software Market in Retail Industry 2008-2012. 2009.
- [15] Emerald Publishing. A novel approach based on utility mining for store layout: a case study in a supermarket. 2017. <https://www.proquest.com/pq1academic/docview/1874315473/3CDF7480B1974F29PQ/1?accountid=32236>
- [16] Gutierrez, L. Universidad de las Americas, Puebla. Reconocimiento de Patrones. 2000. http://catarina.udlap.mx/u_dl_a/tales/documentos/msp/gutierrez_g_l/capitulo2.pdf
- [17] Beltran, M. Minería de Datos. Facultad de Ciencias de la Computación. Universidad Autónoma de Puebla. 2020. <http://bbeltran.cs.buap.mx/NotasMD.pdf>
- [18] Brandesign. Como Predecir y Analizar los Patrones de Compra de Tus Clientes. 2022. <https://www.brandesign.es/marketing/como-predecir-y-analizar-los-patrones-de-compra-de-tus-clientes/>
- [19] Myservname.com. Algoritmo Apriori en Minería de Datos: Implementación con Ejemplos. 2022. <https://es.myservname.com/apriori-algorithm-data-mining>
- [20] Orozco, M. Método de Reglas de Asociación Para el Análisis de Afinidad Entre Objetos de Tipo Texto. 2017. <https://repositorio.cuc.edu.co/bitstream/handle/11323/165/72208501.pdf?sequence=1>
- [21] Amat, J. Reglas de Asociación y Algoritmo a Priori con R. Ciencia de Datos.net. 2018. [https://www.cienciadedatos.net/documentos/43_reglas_de_asociacion#:~:text=El%20algoritmo%20a%20priori%20hace%20una,primero\)%20puede%20ser%20frecuente%2080%209D.](https://www.cienciadedatos.net/documentos/43_reglas_de_asociacion#:~:text=El%20algoritmo%20a%20priori%20hace%20una,primero)%20puede%20ser%20frecuente%2080%209D.)