



FACULTAD DE CIENCIAS ECONÓMICAS Y EMPRESARIALES

**MÁSTER EN GESTIÓN ESTRATÉGICA Y NEGOCIOS INTERNACIONALES**

**ANÁLISIS CESTA DE LA COMPRA: ESTUDIO DEL MÉTODO  
MARKET BASKET ANALYSIS: RESEARCH METHOD**

Trabajo Fin de Máster presentado por Dña. Paula Badía Cambriles, ( [paulabadiacambri-les@hotmail.es](mailto:paulabadiacambri-les@hotmail.es) ), en el curso académico 2018/2019 en la convocatoria ordinaria de Junio de 2019, siendo el tutor del trabajo Prof. Dr. Manuel J. Sánchez Franco.

**Vº. Bº del Tutor:**

**Alumna:**

**D. Manuel J. Sánchez Franco**

**Dña. Paula Badía Cambriles**

Sevilla, Junio 2019.

## **RESUMEN**

En el presente escrito se aborda el estudio de la metodología denominada análisis cesta de la compra, una técnica perteneciente a la minería de datos. Desarrollada en un principio para ser aplicada en el ámbito del marketing, su aplicación ha traspasado a campos de investigación más diversos. La finalidad que se persigue es identificar asociaciones entre diversos grupos de productos o categorías. Además de validar asociaciones de productos que pueden considerarse obvias, debido a su reiterada compra conjunta, el análisis cesta de la compra busca la contingencia entre productos cuya asociación no es evidente.

Es asimismo destacable el uso que se realiza de datos considerados como inservibles y desordenados en la gestión de búsqueda de asociaciones; siendo la principal fuente de información disponible la base de datos que recoge las transacciones que realizan los consumidores en cada compra.

**PALABRAS CLAVE:** Análisis de datos, minería de datos, reglas de asociación, análisis cesta de la compra.

## **ABSTRACT**

The current essay offers an approach to the market basket analysis methodology, known to be a technique to mining data. Once conceived as a technique with an exclusive use on marketing studies, its applications have nowadays moved on to a broader spectrum of investigation fields such as biotechnology. Its main purpose is to identify associations between different groups or categories of products. Aside from obvious associations between those products that are often purchased concurrently, this analysis seeks the contingency across products that have a not-so-defined association at first.

It is key to highlight the way this analytical method exploits apparently unuseful data, building from raw data a structured set of analytical information about customer's transactions on each of their purchases.

**KEYWORDS:** Data analysis, mining data, association rules, market basket analysis.

## ÍNDICE.

1. Introducción.....	6
2. Objetivos.....	6
3. <i>Data warehouse</i> .....	7
4. <i>Business Intelligence</i> .....	7
5. <i>Knowledge Discovery in Databases</i> .....	9
5.1. Procesos.....	9
6. Minería de datos.....	11
6.1. Etapas .....	12
6.2. Tareas.....	15
6.3. Reglas de dependencia.....	16
7. Reglas de asociación.....	17
7.1. Técnica.....	18
7.2. Categóricas vs cuantitativas.....	19
7.3. Algoritmo A priori.....	22
8. Análisis cesta de la compra.....	23
8.1. Relaciones de contingencia.....	28
8.2. Dimensionalidad.....	29
8.3. Heterogeneidad de los consumidores.....	29
9. Caso práctico 1: Cesta de la compra en supermercados.....	30
9.1. Leche entera.....	32
9.2. Verduras.....	33
9.3. Bollería.....	34
9.4. Refresco.....	35
9.5. Yogurt.....	36
9.6. Artículos de higiene.....	37
9.7. Arroz.....	38
9.8. Ternera.....	39
9.9. Huevos.....	40
9.10. Pescado envasado.....	41
9.11. Pan integral.....	42
9.12. Asociaciones según lift.....	43
9.13. Asociaciones según soporte.....	43

9.14. Asociaciones según confianza.....	44
9.15. Técnicas de minería gráfica.....	45
9.16. Técnicas de minería gráfica aplicada al caso práctico 1.....	45
9.17. Conclusión caso práctico 1.....	46
10. Sistemas de recomendación.....	49
10.1. Sistema de recomendación en bibliotecas.....	49
10.2. Técnicas basadas en conocimientos.....	50
11. Caso práctico 2: sistema de recomendación propuesto para la Biblioteca de la Universidad de Sevilla.....	51
11.1. Facultad de Derecho.....	55
11.2. Facultad de Ciencias Económicas y Empresariales.....	58
11.3. Facultad de Ciencias de la Educación.....	61
11.4. Facultad de Turismo y Finanzas.....	62
11.5. Facultad de Psicología.....	65
11.6. Facultad de Filosofía.....	69
12. Conclusiones.....	71
13. Referencias bibliográficas.....	73
Anexo 1.....	76

Cuadro 1: Principales técnicas analíticas utilizadas en el proceso de <i>Business Intelligence</i> .....	8
Figura 1: Esquema del proceso <i>Knowledge Discovery in Databases</i> .....	11
Figura 2: Esquema de pasos que se realizan en la metodología minería de datos.....	15
Tabla 1: Ejemplo de reglas de asociación cuantitativas.....	21
Tabla 2: Clasificación booleana.....	21
Tabla 3: Ítemsets frecuentes extraídos de la muestra.....	21
Figura 3: Modelo esquemático para los datos recogidos del análisis cesta de la compra.....	23
Tabla 4: Tabla de transacciones para ejemplo ilustrativo.....	25
Cuadro 2: Extracto del conjunto de transacciones registradas.....	31
Figura 4: Frecuencia relativa de las categorías de productos.....	32
Tabla 5: Productos más frecuentes en la cesta de la compra de los encuestados.....	32
Tabla 6: Tabla de asociaciones con leche entera en el lado derecho (RHS).....	32

Tabla 7: Tabla de asociaciones con leche entera en el lado izquierdo (LHS).....	33
Tabla 8: Tabla de asociaciones con verduras en el lado derecho (RHS).....	33
Tabla 9: Tabla de asociaciones con verduras en el lado izquierdo (LHS).....	34
Tabla 10: Tabla de asociaciones con bollería en el lado derecho (RHS).....	34
Tabla 11: Tabla de asociaciones con bollería en el lado izquierdo (LHS).....	35
Tabla 12: Tabla de asociaciones con refresco en el lado derecho (RHS).....	35
Tabla 13: Tabla de asociaciones con refresco en el lado izquierdo (LHS).....	36
Tabla 14: Tabla de asociaciones con yogurt en el lado derecho (RHS).....	36
Tabla 15: Tabla de asociaciones con yogurt en el lado izquierdo (LHS).....	36
Tabla 16: Tabla de asociaciones con artículos de higiene en el lado derecho (RHS)....	37
Tabla 17: Tabla de asociaciones con artículos de higiene en el lado izquierdo (LHS).....	37
Tabla 18: Tabla de asociaciones con arroz en el lado derecho (RHS).....	38
Tabla 19: Tabla de asociaciones con arroz en el lado izquierdo (LHS).....	39
Tabla 20: Tabla de asociaciones con ternera en el lado derecho (RHS).....	39
Tabla 21: Tabla de asociaciones con ternera en el lado izquierdo (LHS).....	40
Tabla 22: Tabla de asociaciones con huevos en el lado derecho (RHS).....	40
Tabla 23: Tabla de asociaciones con huevos en el lado izquierdo (LHS).....	40
Tabla 24: Tabla de asociaciones con pescado envasado en el lado derecho (RHS)....	41
Tabla 25: Tabla de asociaciones con pescado envasado en el lado izquierdo (LHS)....	41
Tabla 26: Tabla de asociaciones con pan integral en el lado derecho (RHS).....	42
Tabla 27: Tabla de asociaciones con pan integral en el lado izquierdo (LHS).....	42
Tabla 28: Tabla de asociaciones con lift más significativos.....	43
Tabla 29: Tabla de asociaciones con soporte más significativos.....	44
Tabla 30: Tabla de asociaciones con confianza más significativas.....	44
Figura 5: Asociaciones a través de minería gráfica.....	46
Figura 6: Técnicas de recomendación y sus respectivas fuentes de conocimiento.....	50
Figura 7: % de préstamos per cápita relativizado al caso de la Facultad de Ciencias de la Educación (mayor % de préstamos per cápita).....	52
Figura 8: % diario de préstamos per cápita por centro seleccionado.....	53
Figura 9: % diario de préstamos per cápita por curso y centro seleccionado.....	53
Figura 10: Diagrama de violín (y caja) representando el eje relativizado por centro.....	54

Figura 11: Red de palabras según frecuencia de aparición por bigramas en los títulos de libros prestados.....	54
Figura 12: % de préstamos de los principales títulos de libros por población y centro.....	55
Tabla 31: Tabla de asociaciones de libros pertenecientes a la Biblioteca de la Facultad de Derecho ordenados por el indicador soporte.....	56
Tabla 32 : Tabla de asociaciones de libros pertenecientes a la Biblioteca de la Facultad de Derecho ordenados por el indicador lift.....	57
Tabla 33: Tablas de asociaciones según el curso del alumno y el libro más prestado en la Facultad de Derecho.....	58
Tabla 34: Tabla de asociaciones de libros pertenecientes a la Biblioteca de la Facultad de Ciencias Económicas y Empresariales ordenados por el indicador lift.....	59
Tabla 35: Tabla de asociaciones de libros pertenecientes a la Biblioteca de la Facultad de Ciencias Económicas y Empresariales ordenados por el indicador soporte.....	60
Tabla 36: Tablas de asociaciones según el curso del alumno y el libro más prestado en la Facultad de Ciencias Económicas y Empresariales.....	61
Tabla 37: Tabla de asociaciones de libros pertenecientes a la Biblioteca de la Facultad de Ciencias de la Educación ordenados por el indicador soporte.....	62
Tabla 38: Tabla de asociaciones de libros pertenecientes a la Biblioteca de la Facultad de Turismo y Finanzas ordenados por el indicador lift.....	63
Tabla 39: Tabla de asociaciones de libros pertenecientes a la Biblioteca de la Facultad de Turismo y Finanzas ordenados por el indicador soporte.....	64
Tabla 40: Tablas de asociaciones según el curso del alumno y el libro más prestado en la Facultad de Turismo y Finanzas.....	65
Tabla 41: Tabla de asociaciones de libros pertenecientes a la Biblioteca de la Facultad de Psicología ordenados por el indicador lift.....	66
Tabla 42: Tabla de asociaciones de libros pertenecientes a la Biblioteca de la Facultad de Psicología ordenados por el indicador soporte.....	67
Tabla 43: Tablas de asociaciones según el curso del alumno y el libro más prestado en la Facultad de Psicología.....	68
Tabla 44: Tabla de asociaciones de libros pertenecientes a la Biblioteca de la Facultad de Filosofía ordenados por el indicador lift.....	69
Tabla 45: Tabla de asociaciones de libros pertenecientes a la Biblioteca de la Facultad de Filosofía ordenados por el indicador soporte.....	70
Tabla 46: Tablas de asociaciones según el curso del alumno y el libro más prestado en la Facultad de Filosofía.....	71

## 1. INTRODUCCIÓN

La tecnología que concierne a las bases de datos se ha caracterizado por el cambio experimentado en la búsqueda y desarrollo de actividades desde el año 1980; desplegándose nuevos y poderosos sistemas de bases de datos. Como consecuencia ha habido un desarrollo exponencial de sistemas de hardware y software, traduciéndose este hecho en el desarrollo de equipos informáticos, y sistemas de recogida de datos y almacenaje que permiten dar cobertura a las nuevas necesidades (Raorane, Kulkarni & Jitkar, 2012).

Dichas herramientas promueven aún más la creación de bases de datos por parte de las empresas, desembocando en la generación de información acerca de las industrias como un activo intangible valioso para las organizaciones, generado a través de los datos que se almacenan en repositorios disponibles para la transacción de los mismos y permitiendo analizar dichos datos. Más aún, los sistemas computacionales implementados en las entidades permiten almacenar y procesar una gran cantidad de datos, gracias al desarrollo de complejos modelos y algoritmos que extraen conocimientos de las bases de datos.

No obstante, el uso efectivo de esa cantidad de datos puede llegar a ser un problema. Los usuarios pudieren no saber manejar la gran cantidad de datos o, en su defecto, no encontrar la información que buscan de una forma rápida y eficaz, dando como resultado un conocimiento escaso, a pesar de poseer una base de datos abundante de donde extraerlo (Yi, Chen & Cong, 2018).

Por este motivo, el principal objetivo que concierne a la realización del presente escrito es ofrecer una metodología que permita establecer patrones eficientes entre diversos productos que consientan a las organizaciones optimizar sus estrategias a partir de la información que recogen diariamente de sus clientes.

En particular, y en el marco de la presente investigación se ha considerado que la metodología de análisis basada en la cesta de la compra es idónea para identificar relaciones entre ítems, grupos de productos o categorías. Esta técnica, conocida también como minería de reglas de asociación, permite utilizar aquellos datos que se calificaron en un principio de inservibles y desordenados como datos que ayudan a construir teorías dinámicas, las cuales permiten establecer estrategias de marketing en las organizaciones.

## 2. OBJETIVOS

El principal objetivo que concierne al presente escrito desemboca en la revisión tanto teórica como práctica del método análisis cesta de la compra. Los datos que poseen las empresas tienen una información implícita valiosa que debe ser analizada y ejecutada en forma de toma de decisiones para proporcionar mejores resultados. Por este motivo, el interés del escrito reside en el desarrollo de una revisión teórica desde el almacenamiento de datos hasta la ejecución del análisis de los mismos.

Consecuentemente, para verificar la utilidad que reside en el estudio de los datos pertenecientes a una organización, se realizan dos casos prácticos en los que se aplica la metodología de análisis cesta de la compra, apreciando una serie de asociaciones entre los diversos productos que conforman las bases de datos analizadas, permitiendo así

realizar un conjunto de recomendaciones aplicables al ámbito de la organización, con el fin de mejorar la satisfacción en el proceso de toma de decisión del consumidor y su grado de compromiso y lealtad.

### **3. DATA WAREHOUSE**

El proceso de recogida de datos del consumidor es ejecutado diariamente; por ejemplo, los datos generados por cada compra que realiza un cliente. En este contexto, las organizaciones han implantado una serie de sistemas que permiten analizar la elevada cantidad de datos que adquieren para transformarla en información valiosa, asumiendo que la generación de información es un proceso arduo. Las organizaciones ejecutan y analizan una ingente cantidad de datos que genera a su vez ventajas competitivas en un mercado complejo y dinámico.

La recolección de datos no se limita al simple proceso de obtención de datos de los consumidores, sino que deben ser almacenados y depurados para que la utilización de los mismos se realice de una manera eficaz. Por ello, debe destacarse el concepto denominado *data warehouse*. Se define como un almacén de datos que permite acoger e integrar datos en un mismo espacio, procediendo a su vez a la depuración de su contenido; es decir, se desea que la información obtenida entre los diversos departamentos que constituyen la entidad no se reitere, integrando la información desarrollada por la organización. En definitiva, *data warehouse*, se caracteriza por almacenar los datos recogidos de una forma integrada, heterogénea, autónoma y permitiendo su distribución cuando se precisa (Hüsemann, Lechtenbörger & Vossen, 2000).

En suma, la función principal es pues almacenar la información obtenida por la entidad. Asimismo, permite ordenar la información almacenada para posteriormente extraer los conocimientos requeridos sobre sus usuarios, optimizando de esta forma el valor que se les ofrece en el mercado. Este proceso de depuración de la información genera un historial de datos.

### **4. BUSINESS INTELLIGENCE**

La cantidad de datos que posee una empresa puede llegar a ser de incalculables dimensiones. Las organizaciones mecanizan pues este proceso; la compleja actividad queda alejada de la posibilidad del control humano sin apoyo tecnológico. En este sentido, se debe hacer mención de la metodología *Business Intelligence*. Este concepto hace referencia a un conjunto de procesos y aplicaciones que permiten obtener, almacenar, acceder y analizar datos. *Business Intelligence* permite obtener y almacenar los datos precisos para convertirlos posteriormente en información relevante para las mismas mediante análisis precisos (Roldán, Carrión & González, 2012).

*Business Intelligence* está compuesto por diversas claves, sin embargo, se destacan dos especialmente relevantes para el presente estudio; una de las cuales se ha explicado anteriormente, *data warehouse*, y, la otra hace referencia a las funciones de analítica empresarial. Este segundo componente engloba precisamente una serie de herramientas que permite analizar y operar con los datos que se han recogido y almacenado en el mencionado *data warehouse*. Entre las herramientas a las que se hace mención se destacan las técnicas analíticas predictivas, realzando la importancia de la herramienta

denominada minería de datos. Asimismo, *Business Intelligence* apuesta por la incorporación tanto de modelos matemáticos como de metodologías de análisis, debido a que la aplicación de ambos en una serie de datos abastece a la organización con información para obtener conocimiento y facilitar el proceso de toma de decisiones.

El Cuadro 1 recoge diversas técnicas que se utilizan en *Business Intelligence* para analizar los datos que se han recogido de los consumidores; como se observa, la estadística posee un papel fundamental en todas las técnicas recogidas. En el presente estudio nos centraremos en la primera técnica analítica que aparece en el cuadro, minería de datos. Esta elección se debe a que es un proceso iterativo cuyo progreso se alcanza a través del descubrimiento, siendo la finalidad de esta metodología la búsqueda de nueva, válida y significativa información en las bases de datos de las empresas (Kantardzic, 2011).

**Cuadro 1:** Principales técnicas analíticas utilizadas en el proceso de *Business Intelligence*.

<b>Técnicas analíticas</b>
<i>Minería de datos (incluida la minería de reglas de asociación, clasificación, análisis de clústeres y redes neuronales, entre otras técnicas)</i>
<i>Análisis de anomalías</i>
<i>Análisis espacial y temporal</i>
<i>Análisis de red y minería de gráficos</i>
<i>Minería de textos y lingüística computacional</i>
<i>Minería de opiniones y análisis de sentimiento</i>
<i>Optimización y simulación</i>
<i>Análisis de series de decisiones, logística, previsión y series de tiempo</i>
<i>Econometría</i>

Fuente: Bramer (2007).

En suma, *Business Intelligence* combina la recopilación de datos, el almacenamiento de datos y gestionan el conocimiento a través de herramientas analíticas con la finalidad de ofrecer una información interna compleja que permita planificar y tomar decisiones. *Business Intelligence* se caracteriza por ofrecer información accionable en el momento que se requisa y en el formato adecuado para facilitar la toma de decisiones (Negash, 2004).

Cabe señalar la diferencia existente entre *Business Intelligence* y *Knowledge Discovery in Databases*, que precisamos en el siguiente epígrafe. La finalidad perseguida por *Business Intelligence* es obtener y almacenar datos precisos para obtener información, es decir, descubrir y explicar datos relevantes latentes. Por su parte, *Knowledge Discovery in Databases* hace referencia a los procesos de búsqueda, selección, organización, destilación y comprensión de la información que permite al usuario mejorar su comprensión del área de interés (Herschel & Jones, 2005). La finalidad que persigue *Knowledge Discovery in Databases* es también mejorar el uso de la información y el conocimiento disponible en la organización; no obstante, se ampara en el conocimiento subjetivo, operando de manera eficaz la información no estructurada y el conocimiento tácito (Wang & Wang, 2008). En el siguiente epígrafe se desarrolla la metodología *Knowledge Discovery in Databases*.

## 5. KNOWLEDGE DISCOVERY IN DATABASES

A partir del desarrollo de la tecnología se han podido extraer cada vez más datos de los consumidores; no obstante, la extracción de datos no es suficiente para generar contenido útil que ayude a las empresas a mejorar el desarrollo de su actividad. De esta manera, *Knowledge Discovery in Databases* adquiere un papel fundamental cuando se pretende extraer información para las organizaciones.

El concepto *Knowledge Discovery in Databases* se relaciona con métodos de extracción de conocimiento a través de la base de datos que posee la organización, siendo la única finalidad de dicho método conseguir información relevante a partir de los datos proporcionados. En este sentido, se define la metodología *Knowledge Discovery in Databases* como el proceso general que se aplica para extraer toda la información implícita y útil que reside en los datos que posee la organización, mientras que la minería de datos es la aplicación específica de algoritmos para extraer patrones del conjunto de datos.

En suma, *Knowledge Discovery in Databases* está compuesto por una serie de pasos como son preparar, seleccionar, limpiar los datos, y aportar conocimiento tras haber estudiado los resultados (Fayyad, Piatetsky-Shapiro & Smyth, 1996).

Siendo *Knowledge Discovery in Databases* un conjunto de prácticas de creación, desarrollo y aplicación del conocimiento con la finalidad de mejorar el rendimiento en la organización (Paiva & Gonçalo, 2008); los modelos empleados por el proceso de *Knowledge Discovery in Databases* se encuentran en el marco del conocimiento tácito y explícito (Nonaka & Takeuchi, 1995), en donde se introduce el procesamiento humano de una manera dinámica para conseguir la mejora del rendimiento. Diferenciándose con la metodología *Business Intelligence* en que el proceso de *Knowledge Discovery in Databases* utiliza la información no estructurada y el conocimiento tácito (Marwick, 2001) que no se aborda en la primera metodología. Sin embargo, ambos procesos se ven beneficiados por el uso de la metodología de minería de datos, aunque pertenezca al conjunto de procesos que forman *Business Intelligence*, en *Knowledge Discovery in Databases* es utilizada la metodología minería de datos como herramienta para ampliar el conocimiento, i.e., aplicándola a una base de datos y obteniendo patrones de compra (Herschel & Jones, 2005).

### 5.1. PROCESOS

La metodología *Knowledge Discovery in Databases* engloba una serie de procesos consecutivos que se ejecutan de forma iterativa.

- **Definir el problema.** Los objetivos perseguidos cuando se aplica *Knowledge Discovery in Databases* deben estar identificados; dichos objetivos han de ser definidos de una manera procesable.

- **Procesamiento de los datos.** Incluye los procesos de recopilación de datos, limpieza de datos, selección de datos y transformación de datos.

En primer lugar, con el proceso de recopilación de datos se obtiene los datos necesarios de distintas fuentes, es decir, tanto de recursos internos como externos. Además, se unifica la manera en las que van a ser presentados y se codifica las diferencias que muestran. La finalidad perseguida en este proceso es unir los datos procedentes de distintas

fuentes para generar una fuente homogénea. En segundo lugar, la limpieza de datos permite verificar y resolver conflictos presentados en los datos, como valores atípicos o erróneos, ambigüedad o falta de datos; se utilizan conversiones y diferentes combinaciones para formar nuevos conjuntos de datos. En consecuencia, este proceso acumula un esfuerzo considerable. En tercer lugar, se realiza el proceso de selección de datos, el cual consiste en seleccionar los datos que han sido considerados relevantes para la tarea de análisis obtenidos a partir de una determinada base de datos; es decir, se seleccionan un conjunto o, en su defecto, un subconjunto de variables o muestras de datos a través de los cuales obtener la información buscada. En último lugar, la transformación de datos consiste en consolidar los datos en formas apropiadas para, posteriormente, poder aplicarlos a la minería y así realizar operaciones de resumen o agregación.

- **Minería de datos.** Se considera un proceso esencial, donde son aplicados diversos métodos para obtener patrones de datos. Los patrones obtenidos que sean considerados de interés son representados en particular o dentro de un conjunto de representaciones a través de las tareas que componen la minería de datos, las cuales se explican más adelante en este escrito. Consecuentemente, el usuario podrá utilizar el método de minería de datos de una forma eficaz si se han ejecutado los pasos anteriores de una forma correcta. A pesar de ser una metodología perteneciente al proceso *Business Intelligence*, la minería de datos beneficia también al proceso de *Knowledge Discovery in Database*, como se ha expuesto anteriormente.

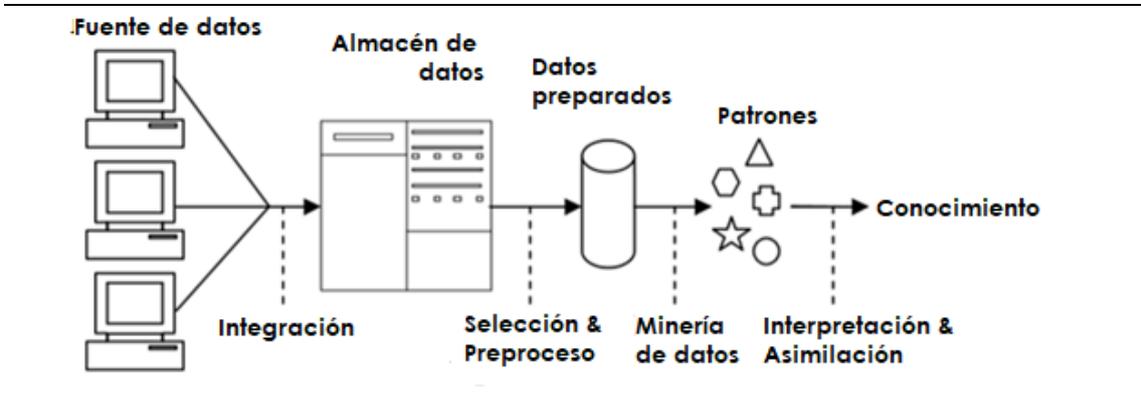
- **Post-minería de datos.** Este proceso está compuesto por evaluación de patrones, desarrollo del modelo, mantenimiento y presentación del conocimiento.

- **Evaluación de patrones.** Se caracteriza por identificar los patrones que se consideran relevantes para representar el conocimiento adquirido, los cuales se basan en diversas medidas, probando la precisión del modelo en un conjunto de datos independiente que no se han utilizado en la creación del modelo. En este sentido, se evalúa la sensibilidad del modelo y se realiza una prueba de su usabilidad. Ejemplificando lo anteriormente expuesto, si se diseña un modelo para predecir la respuesta de un cliente hacia un producto determinado, se podría efectuar una predicción enviando un correo de prueba a un subconjunto de datos para comprobar si las respuestas coinciden con la predicción realizada.
- **Desarrollo del modelo.** El modelo se va a utilizar para predecir resultados en distintos casos. Por lo tanto, se considera que la predicción se usa para alterar el comportamiento organizacional. En consecuencia, la implementación del modelo requiere sistemas informáticos que recojan los datos necesarios para generar una predicción en el momento que se precisa, y así poder tomar decisiones en torno a la predicción realizada.
- **Mantenimiento.** Los modelos que se desarrollan para satisfacer una necesidad de las organizaciones es probable que cambie en el tiempo; esto se debe a que vivimos en un mercado dinámico en constante cambio, que necesita moldear los modelos. Por este motivo, los modelos que se crean requieren de una revalidación constante que permitan evaluar los nuevos datos que se han introducido en el sistema.

En definitiva, el modelo se debe adaptar a los cambios que se realizan en el mercado para que se pueda evaluar las opciones que se le ofrecen al consumidor y así realizar predicciones precisas en todo momento.

- **Presentación del conocimiento.** Son técnicas de visualización y representación del conocimiento que se ha extraído de los datos a través de los modelos que se han aplicado.

**Figura 1:** Esquema del proceso *Knowledge Discovery in Databases*.



Fuente: Chiang, Goes, & Stohr (2012).

Como se refleja en la Figura 1 el proceso de *Knowledge Discovery in Databases* se encuentra constituido por diferentes etapas, entre las cuales podemos observar la metodología de minería de datos. En este contexto, tener una gran cantidad de datos no conlleva conseguir información útil si no se estudian adecuadamente. Por ello, es necesaria la aplicación de herramientas que analicen la gran cantidad de datos que posee la empresa; sin dichas herramientas no se podrá extraer conocimiento para la toma de decisiones. La minería de datos se considera pues un paso esencial en el proceso de *Knowledge Discovery in Databases* (Raorane, Kulkarni & Jitkar, 2012). No obstante, el proceso de *Knowledge Discovery in Databases* no debe ser confundido con la minería de datos. *Knowledge Discovery in Databases* es un proceso que engloba el descubrimiento de conocimiento útil a partir de la base de datos, mientras que la minería de datos es una metodología que se sigue en dicho proceso (Fayyad, Piatetsky-Shapiro & Smyth, 1996).

A continuación, se desarrolla la metodología de minería de datos por su importancia en *Knowledge Discovery in Databases*.

## 6. MINERÍA DE DATOS

Es innegable la utilidad de la estadística como método de explotación. En particular, la minería de datos adquiere un papel clave en el desarrollo del análisis que se debe realizar con la base de datos que posee la organización. Delimitando el concepto de minería de datos, esta se define como un análisis de observación de un conjunto de datos que busca relaciones desapercibidas. La minería de datos ofrece una síntesis de la base de datos perteneciente a la empresa, promoviendo el descubrimiento de relaciones entre bienes (Hand, 2007). Es decir, la minería de datos habilita a las organizaciones a extraer la información que reside de forma implícita en el almacenamiento de datos, i.e., relaciones novedosas no plasmadas anteriormente. Por tanto, busca un grado de primicia que le permita a la entidad desarrollar su actividad organizacional de una

forma más estratégica, evaluando información útil y desconocida para el resto de sus competidores (González & Broitman, 2012).

En este contexto, cabe señalar la importancia de las redes sociales en la era tecnológica. Las redes sociales *online* son servicios webs que permiten a los consumidores desarrollar perfiles públicos dentro de un sistema limitado, generando una conexión entre dichos perfiles de los distintos usuarios (Boyd & Ellison, 2010). De esta forma, las redes sociales poseen un papel muy importante en la generación de datos útiles para las entidades. Introducimos aquí un concepto innovador acorde con las necesidades actuales denominado *mining social media*, basado en la generación de grandes cantidades de información expuestas diariamente en las redes sociales (Gundechea & Liu, 2012).

## 6.1. ETAPAS

A continuación, se enumeran un conjunto de pasos considerados efectivos para la aplicación del proceso de minería de datos.

**Paso 1:** identificar objetivos y metas.

En primer lugar, para poder aplicar el proceso de minería de datos se debe identificar los objetivos y las metas. En este sentido, los objetivos son definidos en un nivel abstracto y menos cuantitativo que las metas, las cuales suelen ser cuantificables. Para entender la diferencia que existe entre ambos conceptos, se señala que un objetivo para una organización pudiere ser aumentar las ventas durante un año fiscal, sin embargo, cuando se identifica una meta se cuantifica, por ejemplo, aumentar las ventas durante el ejercicio fiscal un 15 %, utilizando la misma inversión que el año fiscal anterior.

**Paso 2:** determinar las áreas de negocio y métodos.

Cuando se han definido los objetivos y las metas, se debe determinar sobre qué áreas se van a enfocar y qué métodos se van a utilizar para medir los resultados que se consigan. Por ejemplo, si analizamos de dónde provienen los incrementos del área del marketing pueden que se encuentren en la base de a cuales clientes, de nuevos clientes o de un incremento en las operaciones de marketing.

Por ejemplo, ilustramos el paso 2 con el caso de una campaña publicitaria *online*; los ingresos están determinados por el coste por mil impresiones (CPM). El coste de tener un anuncio expuesto a un millón de impresiones es 1.000 € con un CPM de 1 €; consecuentemente se quiere aumentar el tráfico (en este caso, impresiones) hacia el sitio web y la meta definida es incrementar los ingresos que se generan en un 15 % más los próximos tres meses. Para conseguir dicha meta, se deben analizar las áreas de la organización que pueden actuar. A continuación, se proponen dos acciones de marketing que pueden ser consideradas:

- Incrementar la inversión en marketing de buscadores para dirigir tráfico a la página web.
- Publicitar la página web en otros sitios webs a través de banners.

Una vez que se han considerado diversas opciones, se deberá realizar una evaluación de las ventajas y desventajas que poseen cada una de ellas. Con respecto a las medidas que tienen que ser usadas para medir el resultado que se ha obtenido en el proceso debe ser coherente con la meta que se ha propuesta; es decir, en el ejemplo que se ha

expuesto la meta era incrementar los ingresos un 15 % los tres meses siguientes, por lo que la medida que se utilice para analizar el resultado debe ser apropiada con dicha meta.

**Paso 3:** traducir problemas de negocio en problemas técnicos.

Este proceso resulta fundamental cuando se aplica la metodología de la minería de datos, puesto que una traducción errónea de los problemas de negocio conlleva a una pérdida de recursos y oportunidades. Retomando el ejemplo que se ha expuesto anteriormente, si se quiere anunciar en otra página web, se debe analizar que sitios webs son los más apropiados. Aludiendo a una respuesta técnica a la pregunta dónde se puede anunciar de una forma efectiva, se puede segmentar el tráfico que se pretende generar en una fuente de referencia, los cuales son el sitio de origen de visita del usuario que los redirige a la página que se está anunciando, véase como fuente de referencia Google. En consecuencia, se determinará la eficacia de una fuente de referencia en función del volumen de tráfico, crecimiento del tráfico, perfil del usuario y el coste; todo ello debe ser analizado por una herramienta analítica web que se debe incorporar al proceso de análisis.

**Paso 4:** seleccionar el proceso adecuado de minería de datos y herramientas software.

Las técnicas utilizadas en la metodología de minería de datos se encuentran fundamentada en métodos estadísticos y algoritmos, por ejemplo, regresiones logísticas o árboles de decisión. En este sentido, el software que se emplea en la metodología de la minería de datos implementa técnicas para facilitar al usuario la introducción de los algoritmos.

**Paso 5:** identificar los recursos de datos.

Seleccionados la técnica de minería de datos y el software, los datos adecuados deben ser identificados para hacer un uso eficiente de la minería de datos. A continuación, se ofrece una serie de recursos de datos para su utilización en la minería de datos:

- Recursos internos: base de datos de clientes, base de datos de transacciones, base de datos de historial de marketing, solicitudes telefónicas y de emails, data warehouse.
- Recursos externos: alquiler de datos secundarios, adjuntar datos procedentes de otras entidades.

Identificada la apropiada base de datos que se va a utilizar, es imprescindible preparar la base de datos para su posterior uso, por lo que deberá estar limpia y estandarizada para el análisis que se realizará durante el proceso de la minería de datos.

**Paso 6:** analizar.

Se pueden diferenciar tres etapas a lo largo del análisis de minería de datos: creación del modelo, validación del modelo y verificación del modelo. Estas tres etapas deben ser incluidas en el análisis; en su defecto, si alguna de dichas etapas no es incluida en el proceso existe un riesgo de crear modelos inestables que no proporcionarían ningún resultado valioso para la entidad.

- Creación del modelo. Se extrae un subconjunto de datos para comenzar la construcción del modelo. Habitualmente, se utiliza el 50-70 % de los datos disponibles para realizar este proceso, y los datos que no son utilizados, es decir, los restantes, son empleados en el proceso de validación del modelo.

No obstante, se construye una serie de modelos para, posteriormente, elegir los mejores a través de análisis estadísticos que miden la efectividad de cada uno de ellos.

- Validación del modelo. El subconjunto de datos que no ha sido utilizado en el proceso de creación del modelo se emplea en la validación del modelo que ha sido seleccionado. Se destaca que un error común es emplear el mismo conjunto de datos tanto para construir modelos como para validarlos; por consiguiente, este error desemboca en una falsa inflación del poder que poseen los modelos traduciéndose en la obtención de unos resultados mejores que los que realmente debían ser.

Se resalta la importancia de la validación con una muestra distinta a la utilizada en la construcción del modelo; si funciona correctamente la prueba de validación, se asegura que será efectiva cuando se aplique a situaciones reales.

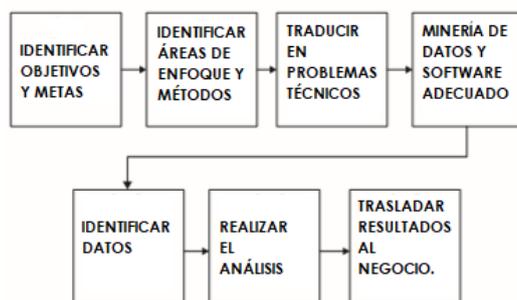
- Verificación del modelo en la realidad. Para comprobar si el modelo que se ha creado *funciona* se recomienda probarlo en una pequeña escala. Por ejemplo, probar un modelo de segmentación, i.e, promociones de marketing, requiere seleccionar un grupo de control y a un grupo de prueba; un grupo de control es una selección aleatoria de todas las posibilidades existentes, mientras que un grupo de prueba está formado por todas las posibilidades que el modelo predice más probables de responder. Por consiguiente, una comparación de las tasas de respuestas entre ambos grupos permitirá dar solidez al modelo que se está analizando. Si el grupo de prueba obtiene una tasa de respuesta más alta que la prueba de control, siendo el resultado estadísticamente significativo, el modelo es sólido, No obstante, si la prueba que se realice falla se tendrá que realizar un análisis minucioso para valorar cuál es el problema que no permite que el modelo funcione cuando se aplica a una situación en la realidad.

**Paso 7:** trasladar resultados analíticos hacia recomendaciones procesables al negocio.

Por último, inferir las recomendaciones de negocios viables aportadas por los resultados obtenidos en el empleo del modelo requiere explicar las principales conclusiones derivadas del análisis en términos que no sean técnicos para que puedan utilizarlas todos los usuarios que precisen de dicha información (Chiu & Tavella, 2008).

En la Figura 2 se recogen la serie de pasos que se han propuesto seguir para conseguir un proceso de minería de datos efectivo.

**Figura 2:** Esquema de pasos que se realizan en la metodología minería de datos.



Fuente: Chiu & Tavellan(2008).

## 6.2. TAREAS

La minería de datos es un proceso de análisis compuesto por una serie de tareas a cumplir y obtener con ello un resultado efectivo para el desarrollo de la actividad organizacional. Las tareas se clasifican en seis: clasificación, estimación, predicción, reglas de asociación, *clustering* y descripción del perfil.

A continuación, se describe cada una de las tareas mencionadas.

- La **clasificación** es considerada una tarea fundamental para realizar el proceso de minería de datos; permite evaluar las características que debe poseer un producto y asignarle un grupo preestablecido. A medida que van entrando nuevos registros se determina el grupo al que pertenecen. Asimismo, se disponen de modelos aplicables a datos no clasificados. Véase el caso del proceso de evaluación de las hipotecas por parte de las entidades bancarias, los demandantes serán clasificados en función del riesgo que presenten y, como consecuencia, se les concederán o no las hipotecas a dichos demandantes.
- La tarea de **estimación** pretende ofrecer un valor para variables continuas desconocidas, como puede ser los ingresos de los consumidores. Por este motivo, las estimaciones de valor para estas variables se basan en variables predictoras. Para ejemplificar esta tarea perteneciente a la minería de datos, proponemos la venta de botas de esquí por parte de una empresa que va a distribuir 500.000 catálogos publicitarios. El problema reside en que se identifican 1,5 millones de esquiadores, sin embargo, al aplicar el método de estimación se envían tales catálogos a los 500.000 esquiadores que más frecuenten las pistas de esquí; el desgaste del producto botas de esquí será mayor y, por lo tanto, la probabilidad de compra a través del catálogo también será mayor.
- La tarea de **predicción** se caracteriza por clasificar los datos en función de comportamientos o estimaciones futuras de valores. Por esta razón, la única manera de evaluar con certeza las decisiones es observando. Además, caben destacar complicaciones añadidas, debido a la relación temporal con las variables o con los predictores que conforman el objetivo. Las herramientas utilizadas para las tareas de clasificación y estimación pueden ser adaptadas para ser aplicadas en la actividad de predicción. Por último, es fundamental basarse en datos históricos para construir un modelo aplicable al comportamiento observado.

Véase el caso de aquellas empresas que quieren predecir la cantidad de consumidores que dejarán de consumir el producto dentro de un año. Para ello, se observarán los datos históricos que posee la empresa y realizarán una estimación del número de clientes que abandonarán su consumo próximamente.

- Las **reglas de asociación** se definen como la agrupación de productos en función de la afinidad que existe entre ellos. Su utilidad reside en identificar oportunidades y diseñar grupos de productos o servicios atractivos para los consumidores. Por ejemplo, un viernes por la noche 1.000 clientes realizaron una compra en un determinado supermercado. De dichos clientes 200 compraron pañales, mientras que 50 compraron cervezas. Si se aplican las reglas de asociación obtenemos la siguiente condición: "Si compran pañales entonces compran cerveza", teniendo un soporte del producto "pañales" de  $200/1.000 = 20 \%$ ; y un soporte de "pañales entonces cerveza" de  $50/1.000 = 5 \%$ . Esta metodología será explicada en profundidad en apartados posteriores del presente escrito.
- El **clustering** se basa en segmentar la heterogeneidad de la población en grupos homogéneos o *clusters*. Esta actividad se diferencia con la tarea de clasificación; los grupos se crean en función de la homogeneidad que presentan. El *clustering* permite segmentar el mercado de una forma eficiente; es decir, en vez de proporcionar una promoción generalizada para todos los clientes, se divide en grupos la base de clientes en función de las características que comparten los consumidores, permitiendo conocer cuál es la promoción que mejor se adapta a cada grupo. Por ejemplo, cuando una campaña de marketing va dirigida a un nicho de mercado específico.
- La tarea de **descripción del perfil** ayuda al proceso de minería de datos a detallar qué ocurre en la base de datos de la empresa, permitiendo aumentar el conocimiento acerca de los productos y consumidores. De esta forma, esta última tarea de minería de datos se considera fundamental para empezar a comprender el comportamiento del consumidor. La tarea de descripción del perfil es utilizada en política para describir el perfil de los votantes, permitiendo conocer por qué en ciertas zonas geográficas de un país se vota más a un partido político que a otro.

### 6.3. REGLAS DE DEPENDENCIA

Como se ha comentado anteriormente, la finalidad del proceso de minería de datos es analizar un conjunto de datos. Para poder realizar dicha actividad es necesario mencionar las reglas de dependencia, las cuales adquieren un papel fundamental en el proceso, puesto que son aplicadas a un conjunto de datos con distinto soporte, que aportarán valor e interés al estudio. En este sentido, las reglas de dependencia son derivadas de un conjunto de elementos caracterizados por no ofrecer el valor esperado, por lo que cumplen con el requisito de suscitar interés que es la característica buscada con este tipo de estudios. Cabe destacar que este rango de elementos se encuentra distribuido de forma no monótona dentro del conjunto.

En particular, sobre la minería por reglas de dependencias recae la tarea de buscar aquellos elementos dependientes que suscitan interés en el estudio. Para ello, hay que desarrollar coeficientes para todo el conjunto; si se centrara únicamente en la obtención de coeficientes para un subconjunto, se perdería una cantidad de información valiosa para la entidad. La finalidad de esta minería por reglas de dependencia consiste

en la definición de un reducido número de patrones, los cuales representen la mayoría de la información alojada en los conjuntos antes mencionados; es decir, se realiza un proceso de simplificación de la información. Debe señalarse que las reglas extraídas a partir de los conjuntos de elementos dependientes gozan de un alto nivel de información; su predicción no es derivada ni esperada. No obstante, una debilidad de la minería por reglas de dependencias reside en cómo medir la información del conjunto de elementos (Savinov, 2004).

Una de las tareas pertenecientes a la minería de datos son las reglas de asociación que identifican dependencias a través de la presencia o ausencia de elementos empleados para generar dichas reglas. En definitiva, las reglas de dependencia son consideradas una generalización de las reglas de asociación, y su aplicación resulta útil para analizar una ingente cantidad de datos de forma eficiente (Silverstein, Brin & Motwani, 1998).

En el siguiente apartado se expone la importancia de las reglas de dependencia para la metodología reglas de asociación, puesto que se busca la dependencia entre productos residentes en una base de datos. Anteriormente ya se ha manifestado la importancia que tienen las seis tareas que forman la minería de datos; sin embargo, hemos optado por centrarnos en un único método de los seis mencionado, las reglas de asociación. Este hecho se fundamenta en que la finalidad que concierne a la minería de datos, como se ha reflejado en apartados anteriores, es extraer información valiosa de una ingente cantidad de datos, dando solución al problema que experimenta las organizaciones ante la situación de tener en desuso una gran cantidad de datos (Agrawal, Imieliński & Swami, 1993).

## **7. REGLAS DE ASOCIACIÓN**

Las reglas de asociación hacen referencia al descubrimiento de relaciones implícitas alojadas en una base de datos, determinando aquellas relaciones que ocurren con una alta frecuencia. Los valores de las variables utilizadas para este estudio se clasifican como binarias; el producto que sea analizado será clasificado como comprado o no comprado. Aplicando las reglas de asociación se evidenciará qué conjunto de productos suelen comprar los clientes.

Las reglas de asociación constituyen una técnica empleada en el campo de la minería de datos que permite descubrir conocimiento en el interior de la base de datos que posee la organización, y ofrece información útil. La aplicación de un análisis estadístico convencional no aportaría el mismo valor que ofrece la información extraída a partir de la aplicación de la técnica de reglas de asociación (Goh & Ang, 2007). Considerada una de las técnicas más importantes en la minería de datos, las reglas de asociación permiten extraer correlaciones que aportan valor, patrones frecuentes o estructuras causales extraídas de la base de datos de la entidad. Por este motivo, es una técnica que es aplicada en diferentes áreas, como pueden ser en el área de biología, informática o marketing (Kotsiantis & Kanellopoulos, 2006).

## 7.1. TÉCNICA

Su aplicación en el dominio del marketing proporciona resultados útiles, los cuales permiten describir grupos de productos, e.g. cómo de factible son diversas uniones de productos para los consumidores. Por este motivo, se ha considerado conveniente profundizar en esta técnica, resumida del siguiente modo: "si un cliente compra X, también comprará Y".

Las reglas de asociación son expresadas como  $X \rightarrow Y$ , donde X e Y son conjuntos de ítems. La interpretación que se extrae de esta fórmula reside en que las transacciones de la base de datos contenidas en X tenderán a estar, también, contenidas en Y. En particular, siendo  $I = \{i_1, i_2, \dots, i_m\}$  ítems y  $D$  una serie de transacciones, donde  $T$  es una transacción, se deduce que  $T \subseteq I$ , es decir,  $I = \{i_1, i_2, \dots, i_m\}$  es un conjunto de atributos que tendrán que ser recalificados por el dominio binario  $\{0,1\}$ . La secuencia de valores  $T$  pertenecientes a  $D$  será generada por la identificación de los atributos con valor 1; formando un identificador único  $TID$ . Un conjunto de artículos  $X \subset I$  es conocido como ítemset. Por lo tanto, una transacción  $T$  contiene  $X$  ítemset, si  $X \subseteq T$ . Siendo una regla de asociación una implicación de la forma  $X \rightarrow Y$ ,  $X \subset I$ ,  $Y \subset I$  y  $X \cap Y = \emptyset$ .

Los principales indicadores de análisis son los siguientes:

La condición  $X \rightarrow Y$  se sustenta en la relación de transacciones  $D$  con confianza  $c$  si  $c$  % de las transacciones en  $D$  contiene tanto  $X$  como  $Y$ .

La regla  $X \rightarrow Y$  posee un soporte  $s$  en el conjunto de transacciones  $D$  si  $s$  % de las transacciones en  $D$  recogen la condición  $X \cap Y$ .

Los datos negativos o los ítems perdidos, no se van a considerar en este estudio puesto que carecen de interés para el mismo (Agrawal, Mannila, Srikant, Toivonen & Verkamo, 1996).

De esta manera, nos encontramos ante dos indicadores relevantes para el presente estudio, confianza y soporte; estos dos indicadores ayudan al investigador que los aplica a conocer la usabilidad que posee la aplicación de la regla de asociación en el análisis que esté realizando.

La infraestructura soporte-confianza es pues aplicada para conocer qué tipo de dependencia se presenta en los ítems de una determinada base de datos. A pesar de los beneficios que aporta al análisis dichos indicadores (soporte y confianza), puesto que miden la incertidumbre de una regla de asociación, el soporte se encuentra limitado en su función, debido a que no proporciona un *feedback* adecuado por su condición de representación a partir de las transacciones, en lugar de representarlo a través del número de ítems que lo conforman. Por este motivo, se han desarrollado diferentes medidas sobre soporte y confianza que desembocan en diferentes modelos para aplicarlos a la minería por reglas de asociación, los cuales se exponen a continuación.

En primer lugar, Piatetsky-Shapiro razona que la regla  $X \rightarrow Y$  no se puede considerar interesante si:

$$\text{soporte}(X \rightarrow Y) \approx \text{soporte}(X) \times \text{soporte}(Y)$$

Acorde con la interpretación de probabilidad soporte  $(X \cap Y) = p(X \cap Y)$  y confianza  $(X \rightarrow Y) = p(Y | X) = p(X \cap Y) / p(X)$ . Por lo tanto, el argumento aportado por Piatetsky-Shapiro se interpreta como:

$$p(X \cap Y) \approx p(X) p(Y)$$

En este sentido,  $X \rightarrow Y$  no puede utilizarse como regla si  $p(X \cap Y) \approx p(X) p(Y)$ . La explicación de este suceso reside en la teoría de la probabilidad, ya que  $(X \cap Y) \approx p(X) p(Y)$  se interpreta como X es aproximadamente independiente de Y.

En segundo lugar, una definición estadística del término dependencia para los conjuntos X e Y es:

$$\text{Lift}(X, Y) = \frac{p(X \cap Y)}{p(X)p(Y)}$$

Esta expresión mide la incertidumbre de las reglas de asociación; refiriéndose al interés que reside en Y una vez dado X. Por este motivo, el valor obtenido a partir de esta fórmula tendrá más dependencia cuanto más alejado esté de 1. De este modo nos encontramos con tres posibles casos:

$p(X \cap Y) / (p(X) p(Y)) = 1$ , X e Y son independientes.

$p(X \cap Y) / (p(X) p(Y)) > 1$ , Y es positivamente dependiente de X.

$p(X \cap Y) / (p(X) p(Y)) < 1$ , Y es negativamente dependiente de X.

Haciendo referencia a la infraestructura de soporte-confianza se deberán cumplir pues las siguientes tres condiciones de mínimo soporte, mínima confianza y mínimo lift para que se pueda extraer de la regla  $X \rightarrow Y$  una regla de interés válida (Zhang & Zhang, 2002).

Entonces, siendo I un conjunto de ítems en la database D; X, Y ( $\subseteq I$ ) son conjuntos de ítems;  $X \cap Y = \emptyset$ ,  $p(X) = \emptyset$  y  $p(Y) = \emptyset$ . Los umbrales que deben cumplirse son: mínimo soporte, mínima confianza y mínimo lift (mínimo lift > 0), estos valores suelen ser dados por los expertos.

(1)  $p(X \cap Y) \geq \text{mínimo soporte}$ . Si el número de intervalos es grande, el soporte podría ser bajo. Por este motivo, en el estudio se limita el soporte que se quiere analizar.

(2)  $p(Y | X) \geq \text{mínima confianza}$ . La mínima confianza se utiliza, puesto que existe cierta información que se pierde cuando se forman los intervalos. La información que se pierde aumenta en función de que el tamaño del intervalo se agrande.

(3)  $\left| \frac{p(X \cap Y)}{p(X)p(Y)} - 1 \right| \geq \text{mínimo lift}$

## 7.2. CATEGÓRICAS VS CUANTITATIVAS

Las reglas de asociación están compuestas por un lado izquierdo (*left-hand side*, LHS) y por un lado derecho (*right-hand side*, RHS). La explicación de esta diferenciación

desemboca en que el lado izquierdo de la regla de asociación (LHS) alberga una descripción perteneciente a un subconjunto de la base de datos, siendo el lado derecho (RHS) de la regla la descripción de un comportamiento imperante en dicho subconjunto. Lo anterior hace referencia a una estructura general que se aplica a diversos tipos de reglas. No obstante, se ejemplifica a continuación la estructura con un tipo concreto de regla: de categóricas a cuantitativas, aplicando un número ilimitado de atributos a ambos lados.

Antes de comenzar a desarrollar una regla de categóricas a cuantitativas se debe conocer que  $E = \{e_1, \dots, e_m\}$  es un conjunto de atributos para una base de datos  $D$  de transacciones. Siendo  $E_Q \subseteq E$  el conjunto de atributos cuantitativos, mientras que  $E_C \subseteq E$  es el conjunto de atributos categóricos, y  $C$  el conjunto de todos los posibles valores categóricos. Cada transacción en  $D$  es un conjunto  $t = \{\langle e_1, v_1 \rangle, \dots, \langle e_m, v_m \rangle\}$  de  $m$  atributos con sus correspondientes valores; para cada  $i$ , si  $e_i$  es categórica  $v_i \in C$ , sin embargo, si  $e_i$  es cuantitativa es  $v_i \in \mathbb{R}$ .

Por consiguiente, el lado izquierdo (LHS) representa una serie de atributos categóricos, mientras que el lado derecho (RHS) hace referencia a un vector de valores principales para varios conjuntos de atributos cuantitativos.

En primer lugar, el lado izquierdo (LHS) de la regla es una sucesión  $X \subseteq E_C \times C$  de atributos categóricos junto con sus correspondientes valores categóricos. Una serie  $X$ , denominada el perfil, referencia a un subconjunto de la base de datos. Por lo tanto, para una transacción  $t = \{\langle e_1, v_1 \rangle, \dots, \langle e_m, v_m \rangle\}$ ,  $t$  poseerá el perfil  $X$  si  $X \subseteq t$ ;  $T_X$  hace referencia al conjunto de transacciones con perfil  $X$ . En segundo lugar, el lado derecho (RHS) consta de un vector de valores medios recogidos de un grupo de atributos cuantitativos, con la media extraída de las transacciones que coinciden con el perfil del lado izquierdo (LHS). Asimismo, para una serie de atributos cuantitativos  $J$ , y un conjunto de transacciones  $T \subseteq D$ , se deduce que  $Media_J(T)$  es el vector de los valores medios de los atributos en  $J$  para el conjunto  $T$ . En el caso del lado derecho (RHS) obtenemos que  $Media_J(T_X)$  para ciertas  $J \subseteq E_Q$ .

La finalidad de aplicación de la presente regla es que dicha regla registra un comportamiento extraordinario si la media de los atributos en  $J$  sobre las transacciones  $T_X$  es significativamente diferente al resto; para ello, se deberá realizar una comparación entre la media  $T_X$  y la media del complemento  $D - T_X$ . De esta forma, podemos decir que  $Media_J(T_X)$  es diferente a  $Media_J(T_Y)$ , siendo  $Media_J(T_X) \neq Mean_J(T_Y)$ , y pudiendo inferir estadísticamente que para cada  $e \in J$  la media del atributo  $e$  en  $T_X$  y  $T_Y$  son diferentes.

En definitiva, una regla de asociación de categórica a cuantitativa se define como una regla con la forma  $X \Rightarrow Media_J(T_X)$ , entendiéndose que  $X$  es un perfil de atributos categóricos ( $X \subseteq E_C \times C$ ),  $J$  es una serie de atributos cuantitativos ( $J \subseteq E_Q$ ) y, por lo tanto,  $Media_J(T_X) \neq Media_J(D - T_X)$  (Aumann & Lindell, 2003)

Para ejemplificar lo expuesto anteriormente, se propone un estudio compuesto por dos atributos cuantitativos (edad y número de automóviles) y un atributo categórico (estado civil: casado).

**Tabla 1:** Ejemplo de reglas de asociación cuantitativas

ID	Edad	Casado	Nº automóviles
100	23	No	1
200	25	Sí	1
300	29	No	0
400	34	Sí	2
500	38	Sí	2

Fuente: Srikant & Agrawal (1996).

La Tabla 1 recoge un ejemplo de reglas de asociación cuantitativas. Para poder analizar la muestra se realizarán dos intervalos de edad (20,...,29) y (30,...,39); una vez realizados dichos intervalos se aplicará la clasificación booleana asignando el valor "1" al atributo que le corresponda (atributo1, valor1), siendo "0" el valor asignado al caso contrario.

Como se observa en la tabla 2 se ha aplicado la clasificación booleana, la cual permitirá conocer las reglas de asociación cuantitativas que se busca con este ejemplo.

La Tabla 3 muestra los resultados obtenidos, los cuales aparecen con más asiduidad aplicando un soporte mínimo = 40 % y una confianza mínima = 50 % (Srikant & Agrawal, 1996).

**Tabla 2:** Clasificación booleana.

ID	Edad	Edad	Casado:	Casado:	Nº	Nº	Nº
	20,...,29	30,...,39	Sí	No	automóviles:	automóviles:	automóviles:
					0	1	2
100	1	0	0	1	0	1	0
200	1	0	1	0	0	1	0
300	1	0	0	1	1	0	0
400	0	1	1	0	0	0	1
500	0	1	1	0	0	0	1

Fuente: Srikant & Agrawal, (1996).

**Tabla 3:** Ítemsets frecuentes extraídos de la muestra.

Ítemset	Frecuencia
{{Edad: 20,...,29}}	3
{{Edad: 30,...,39}}	2
{{Casado: Sí}}	3
{{Casado: No}}	2
{{Nº automóviles: 0...1}}	3
{{Edad: 30,...,39} , {Casado: Sí}}	2

Fuente: Srikant & Agrawal (1996).

### 7.3. ALGORITMO A PRIORI

A continuación, se destaca la relevancia del algoritmo denominado *A priori* en la minería por reglas de asociación. Es un algoritmo que permite generar reglas de asociación. Para poder aplicar dicho algoritmo las variables deben ser categóricas y se discretizan las variables numéricas, siendo las variantes más sofisticadas liberadas de esta restricción. En este sentido, el proceso que realiza se basa en el conocimiento adquirido a través de las propiedades repetidas en un conjunto de ítems, denominando estas propiedades como propiedad *A priori*. De este modo, se considerará como elemento frecuente aquel que satisfaga un soporte mínimo para la aplicación del algoritmo *A priori*, y estableciendo el presente algoritmo en todos los conjuntos o subconjuntos que cumplan la condición.

Asimismo, el algoritmo *A priori* se define como iterativo; cada iteración  $i$  genera una serie de elementos candidatos  $C_i$  obtenidos a partir de la base de datos de transacciones para, posteriormente, comprobar si son frecuentes. Aquellos candidatos que presenten mayor frecuencia ( $L_i$ ) serán utilizados para formar el conjunto de candidatos  $C_{i+1}$  para la sucesiva iteración. De este modo, para garantizar una generación de candidatos  $C_{i+1}$  se realizan uniones de conjuntos frecuente  $L_i$  hallados en la interacción anterior; por lo que si poseen  $i+1$  elementos comunes se producirá una unión entre dos conjuntos de elementos; la duplicidad de los elementos integrados en estas uniones será desechada una vez finalice el proceso de unión de las series de elementos. Por este motivo, el proceso se detiene cuando los candidatos contenidos en  $C_{i+1}$  no son frecuentes en la siguiente iteración (Goh & Ang, 2007).

A pesar del atractivo que recae sobre las reglas de asociación, debido a la utilidad que aporta al estudio, se debe resaltar la problemática asociada. La problemática puede ser dividida en tres tipos de reglas que son producidas cuando son aplicadas las reglas de asociación, estos tres tipos de reglas se denominan: procesable, trivial e inexplicable (Berry & Linoff, 2004). Se desarrollan a continuación.

Las reglas procesables se caracterizan por poseer información de alta calidad. Una vez se localiza el patrón se puede justificar a través de la comprensión de una historia. Véase el caso que propuso un estudio de Wal-Mart el cual indicaba que existía una relación de compra entre muñecas Barbie y chokolatinas. La explicación de este hecho anida en que existen circunstancias que se repiten con este tipo de compra, como puede ser que unos padres compren una muñeca a su hijo por su cumpleaños y, cuyo hermano tenga celos, por lo que para calmar su estado anímico le compran una chokolatina o, también, puede darse el caso que los padres compren la chokolatina, puesto que la actividad de compra ha sido agotadora y la toman como un símbolo de recompensa por el arduo trabajo realizado. Consideremos que esta información recoge un gran interés para las empresas, puesto que les abastece de información para gestionar el punto de venta.

Las reglas triviales se encuentran caracterizadas por ser predecibles, es decir, si la empresa posee una promoción en sus productos, lo más probable es que en el estudio aparezcan juntos dichos productos. Por este motivo, las excepciones que aparezcan en dicho estudio serán susceptibles de análisis, puesto que generan una información valiosa para la empresa.

Las reglas inexplicables son aquellas que hacen referencia a las compras que carecen de aparente explicación.

## 8. ANÁLISIS CESTA DE LA COMPRA

La búsqueda de asociaciones entre los productos que adquieren los consumidores abastece a las empresas de información útil para desarrollar su actividad empresarial. Por esta razón, se introduce en el presente estudio la técnica análisis cesta de la compra (*Market Business Analysis*), debido a que es una técnica perteneciente al campo de la minería de datos, la cual genera asociaciones entre los datos que se encuentran en la base de datos (Raorane, Kulkarni & Jitkar, 2012).

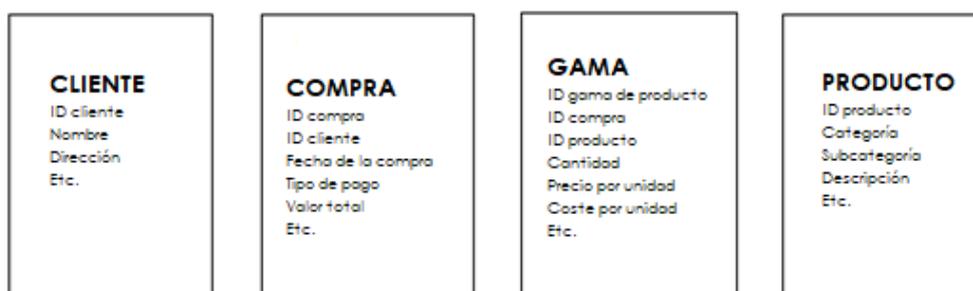
La finalidad que engloba el proceso de análisis de cesta de la compra descansa en la identificación de relaciones entre grupos de bienes o sus categorías. A pesar de utilizarse dicha metodología en diversos campos de estudio, véase bioinformática, ciencia nuclear o geografía, el análisis de cesta de la compra fue específicamente desarrollado para ejecutarlo en el ámbito del marketing (Aguinis, Forcum & Joo, 2013).

El análisis cesta de la compra es una metodología que se desarrolla en el ámbito de la gestión, la cual ilustra y explica la combinación que se debe seguir para obtener avances teóricos en esta materia y, como consecuencia, aplicaciones prácticas significativas. El análisis de cesta de la compra ofrece asociaciones de productos que no son obvias y, por lo tanto, suelen estar encubiertas si se realiza un estudio superficial, es decir, sin profundizar en la cuestión de las asociaciones. Por este motivo, la metodología mencionada permite identificar aquellos productos que son elegidos simultáneamente por los consumidores en una base frecuente y evaluando la extensión con las que son elegidos (Aguinis, Forcum & Joo, 2013).

La metodología análisis cesta de la compra es compleja, debido a que engloba una serie de problemas relacionados con el estudio de los datos recogidos en las transacciones realizadas en los puntos de venta. Asimismo, dichos datos de transacción modelan tres entidades distintas, siendo imprescindibles para la metodología que se está describiendo, las cuales son:

- Clientes
- Compras
- Productos

**Figura 3:** Modelo esquemático para los datos recogidos del análisis cesta de la compra.



Fuente: Berry & Linoff (2004).

El orden es la estructura fundamental que guía el análisis de cesta de la compra, como se observa en la Figura 3. En cualquier caso, si bien un proceso de compra se considera un evento único que realiza un determinado cliente en un determinado momento de tiempo, esta condición no discrimina que dicho cliente compre un conjunto de productos o, en su defecto, un único producto, siendo el canal de compra *online* u *offline*. Dicha transacción está compuesta tanto de la cantidad total de compra como de cualquier otro dato relevante para el estudio; generando un identificador único. Los datos que recoge dicho identificador son pues enlazados con una tabla de referencia que genera una información más descriptiva de cada producto (Berry & Linoff, 2004).

Conviene precisar distintos elementos clave. En primer lugar, a pesar de ser el análisis de cesta de la compra una técnica que tradicionalmente se ha basado en utilizar datos de archivo, también emplea datos primarios. Para depurar dichos datos se utilizan dos reglas de asociación que se han explicado anteriormente, las reglas de asociación categóricas y las reglas de asociación cuantitativas. Las reglas categóricas, las cuales se basan en datos binarios, son las más usadas en la metodología análisis de cesta de la compra, puesto que se desarrolló con la finalidad de analizar la cesta de la compra respaldándose en una forma de clasificación binaria, es decir, si un comprado compraba o no un producto específico. Sin embargo, se desarrolló la posibilidad de introducir una variable continua. La diferencia reside en que las reglas de asociación cuantitativas relacionan una variable categórica con una síntesis estadística perteneciente a una variable continua, mientras que las reglas de asociación categóricas solo permiten relacionar un valor de una variable categórica con el valor de otra variable categórica.

En segundo lugar, el análisis cesta de la compra requiere que la información extraída de los datos no resulte obvia para los investigadores, y aportando lógica a las asociaciones de productos que se busca con tal análisis. Por este motivo, son remarcables tres indicadores fundamentales para comprender la presencia, fuerza y naturaleza de una regla de asociación, los cuales son: *lift*, *soporte* y *confianza*. Su utilización se basa en la complementariedad e información no redundante que aportan al estudio.

Inicialmente se obtiene el índice *lift*. Este índice determina si existe asociación entre los productos, del mismo modo que si dicha asociación es positiva o negativa. El índice *lift* se calcula del modo siguiente:

$$\frac{P(A \cap B)}{P(A) * P(B)}$$

El denominador ilustra el escenario en que los elementos A y B son independientes entre sí. En el numerado se indica que los elementos A y B ocurren simultáneamente. En este sentido, cuando el indicador *lift* es cercano al valor 1, debido a que el numerador y denominador son similares, se deduce que la relación que existe entre el elemento A y el elemento B es de interdependencia. Asimismo, cuanto mayor sea el indicador se traduce en que la presencia del elemento A está asociada con la presencia del elemento B, siendo positiva la relación existente entre ambos elementos. Sin embargo, cuando es inferior a 1 indica que la relación existente entre los elementos A y B es negativa; la presencia de A se encuentra asociada a la ausencia de B.

Para comprender mejor este indicador incluido en el análisis de cesta de la compra se propone un ejemplo. Se propone que en un supermercado se han registrado un total de 10 000 transacciones de compra; para este ejemplo se escogen 4 productos de las

transacciones de compra que han realizado los consumidores: refresco (A), zumo de naranja (B), crema desnatada (C) y donuts (D) (Aguinis, Forcum & Joo, 2013).

**Tabla 4:** Tabla de transacciones para ejemplo ilustrativo.

TRANSACCIÓN	REFRESCO	ZUMO DE NARANJA	CREMA DESNATADA	DONUTS
1	1	1	1	1
2	1	1	1	1
3	1	1	1	1
...				
10	1	1	1	1
11	1	1	1	0
12	1	1	1	0
13	1	1	1	0
...				
30	1	1	1	0
31	1	0	1	0
32	1	0	1	0
33	1	0	1	0
34	0	1	1	0
35	0	1	1	0
...				
500	0	1	1	0
501	0	0	0	1
502	0	0	0	1
503	0	0	0	1
...				
990	0	0	0	1
991	0	0	0	0
...				
10,000	0	0	0	0

Fuente: Aguinis, Forcum & Joo (2013).

La Tabla 4 recoge las transacciones realizadas por los clientes. El valor 1 significa que el consumidor ha adquirido dicho producto, en su defecto, cuando aparece el valor 0 significa que el cliente ha decidido no comprar el producto.

Como refleja la Tabla 4 en las primeras 30 transacciones, el elemento A y el elemento B son comprados simultáneamente. Sin embargo, el producto A (refresco) es adquirido en 33 transacciones, mientras que el producto B (zumo de naranja) es comprado en 497 transacciones. Aplicando el índice lift obtenemos:

$$\text{Lift} = \frac{30/10.000}{(33/10.000) \cdot (497/10.000)} = 18,29$$

Al ser el resultado del índice lift superior a 1, se puede concluir que el producto A (refresco) y el producto B (zumo de naranja) poseen una relación positiva.

Para la siguiente ejemplificación tomamos el producto C (crema desnatada) y el producto D (donuts), ambos productos aparecen comprados simultáneamente en 10 compras del total de 10.000 transacciones, siendo 500 transacciones para el producto C y 500 transacciones para el producto D. Si aplicamos el índice lift obtenemos:

$$\text{Lift} = \frac{10/10.000}{(500/10.000) \cdot (500/10.000)} = 0,4$$

El resultado obtenido es 0,4, siendo el valor inferior a 1, se determina que la relación entre el producto C (crema desnatada) y el producto D (donuts) es negativa, es decir, si el comprador adquiere el producto C existe menos probabilidad de comprar también el producto D.

Resaltando la importancia de inferir las redes de relaciones entre diferentes productos, se destacan dos tipos de relaciones importantes para adquirir conocimiento en este tipo de análisis: productos sustitutivos y productos complementarios. Por un lado, los productos sustitutivos son aquellos que son reemplazables uno por otros, por ejemplo, reemplazar café por té, puesto que ambos satisfacen la misma necesidad; por otro lado, los productos complementarios se adquieren de forma conjunta para satisfacer una necesidad, véase la compra de una camiseta y unos vaqueros que suplen la necesidad de vestirse (McAuley, Pandey & Leskovec, 2015). Analizar las relaciones que poseen los productos entre sí permitirá mejorar la demanda a través de la variable precio, es decir, la reducción del precio de un producto aumenta su demanda, pero también aumentará la demanda del producto complementario del mismo, por tanto, las demandas de ambos productos están asociadas positivamente. Por el contrario, si los productos son sustitutivos, la relación que poseen es negativa, traduciéndose en que la reducción del precio de un producto disminuirá la demanda de su producto sustitutivo.

Aunque el indicador lift proporciona información acerca de la presencia de una asociación, no evalúa estadísticamente la relevancia de la dependencia demostrada entre los productos. Por ofrecer una información reducida se utilizan otros indicadores que complementen la información aportada.

El segundo indicador que se analiza es el soporte. Se representa como  $P(A \cap B)$ . Al ser una probabilidad estará comprendida entre 0 y 1 (o 0 y 100 %). El significado del indicador soporte hace referencia a la probabilidad de que A y B ocurran simultáneamente. Dicha definición se parece al concepto de regla de asociación.

Si se retoma el ejemplo que se ilustra en la tabla 8, se calcula el indicador soporte de la relación positiva entre el producto A (refresco) y el producto B (zumo de naranja):

$$\text{Soporte} = P(A \cap B) = 30 / 10.000 = 3 \%$$

Y, el indicador soporte para la relación negativa existente entre los productos C (crema desnatada) y D (donuts) sería:

$$\text{Soporte} = P(A \cap B) = 10 / 10.000 = 1 \%$$

Se puede apreciar como el valor del indicador soporte posee una gran semejanza entre las dos asociaciones, a pesar de que la primera relación sea positiva y la segunda, en su defecto, sea negativa.

Como se puede comprobar en el ejemplo, una desventaja del indicador soporte reside en que la usabilidad de dicho indicador disminuye a medida que el tamaño y la riqueza de la base de datos es mayor; en dichas situaciones, los valores de este indicador tenderán a ser bajos. Como consecuencia, los valores de soporte similares determinarán una menor utilidad en la distinción de la fuerza de cada regla de asociación analizadas.

Por último, el indicador confianza es la probabilidad de que un conjunto de ítems sea seleccionado habiendo elegido otro conjunto de ítems. El indicador confianza se define como:

$$\text{Confianza} = \frac{P(A \cap B)}{P(A)}$$

Asimismo, este indicador no es influido por el tamaño de la base de datos ni tampoco por la riqueza de esta, al contrario que el indicador soporte; la finalidad de este indicador, como se ha expuesto antes, es conocer la probabilidad de seleccionar un producto, habiendo elegido otro producto anteriormente. Como consecuencia, aunque el número de reglas de asociación no difieran notablemente en valores del indicador soporte, probablemente difieran en los valores proporcionados por el indicador confianza. Cabe resaltar la ventaja que proporciona este indicador, puesto que sobre él reside la capacidad de construir teorías causales; derivándose del hecho que puede ser calculado por dos vías, por un lado, la probabilidad de A dada B y, por otro lado, la probabilidad de B dada A; los resultados de ambas probabilidades pueden ser significativamente diferentes.

Para ejemplificar este indicador, se continua con el ejemplo ilustrado en la Tabla 4. En este sentido, para la asociación del producto "refresco" (A) con el producto "zumo de naranja" (B), existen dos maneras de calcular el presente indicador:

La probabilidad de adquirir zumo de naranja (B) dado que también se ha comprado refresco (A):

$$\text{Confianza} = \frac{30/10.000}{33/10.000} = 91 \%$$

La probabilidad de adquirir refresco (A) dado que también se ha comprado zumo de naranja (B):

$$\text{Confianza} = \frac{30/10.000}{497/10.000} = 6 \%$$

El resultado reflejado aporta una base para formular la hipótesis que concierne que la compra de refresco causa una elevada inclinación hacia la compra de zumo de naranja. Esta posibilidad permite replantear el modo de venta de ambos productos, sugiriendo una posible promoción de refrescos para potenciar la compra de zumos de naranja (Aguinis, Forcum & Joo, 2013).

El proceso de obtener interacciones entre los diversos productos es complejo, puesto que se producen complicadas relaciones de marketing que conciernen a precios, promociones, surtido, espacio y localización de cada unidad de producto que oferta la

organización. Sin embargo, las decisiones relativas a cada unidad de producto no solo afectan a su propia venta, sino que se extiende a la venta de otros productos.

Cabe destacar que existen distintos tipos de interrelaciones. En este caso, se distingue entre las interrelaciones de marketing (efecto cruzado) y las no interrelaciones de marketing (coincidencia). Las interrelaciones de marketing se relacionan con un conjunto de ítems cuya decisión de marketing corresponde a un solo producto, sin embargo, afecta en la compra de otra serie de productos; un ejemplo sería la compra de un detergente y de un suavizante para lavar la ropa. Se detectan las interrelaciones de marketing a través de la estimación de precios cruzados de la demanda, por ejemplo. Se señala además que dichos precios cruzados varían en función del tipo de producto. Por otro lado, el segundo tipo de interrelaciones (coincidencia) no se encuentran determinadas por acciones de marketing. Véase el coste fijo que supone visitar una tienda (Musalem, Aburto & Bosch, 2018).

En este sentido, los sistemas de recomendación muestran a los clientes una enorme variedad de productos con impredecibles oportunidades de conocer sus necesidades, permitiendo a los consumidores realizar asociaciones entre la gran cantidad de productos que ofertan las entidades y dar a conocer nuevos, formando pues una experiencia de valor para los clientes. En este sentido, los modelos explicativos de comportamiento de elección multicategoría ofrecen de manera explícita las dependencias de compra en dicha categoría.

Por consiguiente, en la metodología de análisis de cesta de la compra se pueden diferenciar dos tipos modelos que son utilizados para realizar investigaciones, por un lado, los modelos exploratorios y, por otro lado, los modelos explicativos. Cuando se aplican los modelos exploratorios la finalidad buscada es descubrir diversas interrelaciones de categorías cruzadas basadas en patrones de productos o categorías que han sido adquiridos conjuntamente; este proceso es conocido como análisis de afinidad. Se examinan los efectos de compras de categorías cruzadas en el nivel agregado de la demanda. Por su parte, los modelos explicativos o predictivos de elección multicategoría se centran en la estimación de los efectos de las variables del marketing-mix y se computan explícitamente las dependencias. Se resuelven a través de modelos logit (regresión que obtiene una variable categórica en función de variables independientes) y modelos probit (regresión cuya variable dependiente es bivariable) (Boztuğ & Reutterer, 2008).

### **8.1. RELACIONES DE CONTIGENCIA**

A lo largo del presente escrito se ha manifestado que las relaciones directas son fáciles de identificar siendo importantes para el desarrollo de la actividad empresarial, sin embargo, no recogen toda su complejidad. En este sentido, el análisis cesta de la compra es utilizado para descubrir relaciones de contingencia conocidas como efectos de moderación o interacción. Por este motivo, la contingencia debe ser apreciada como una estrategia que orienta o sugiere maneras en que un fenómeno debe ser conceptualizado y explicado (Boyd, Takacs Haynes, Hitt, Bergh & Ketchen Jr, 2012). Consecuentemente, el análisis cesta de la compra aporta cuál es la fuerza que posee cada asociación de productos, al igual que el alcance de dicha relación en función de diferentes contextos.

La contingencia es pues aproximada en función de varios factores como pueden ser el tiempo y el lugar de la compra. Añadir información contextual al análisis de cesta de la compra permite pues identificar patrones de compra; considerar si dichos patrones de compras ocurren solo en ciertos días señalados o en determinadas zonas geográficas (Tang, Chen & Hu, 2008).

Es destacable el uso de regresiones múltiples para evaluar las relaciones de contingencia en aquellas asociaciones cuyas suposiciones son insostenibles; es decir, se utilizará una regresión múltiple para analizar los efectos de las variables moderadoras categóricas, las cuales muestran las diferencias que se presentan en los distintos grupos (Aguinis, Beaty, Boik & Pierce, 2005).

## **8.2. DIMENSIONALIDAD**

Las organizaciones suelen ofertar una gran cantidad de ítems, en consecuencia, se analizan una infinidad de asociaciones; la dimensionalidad puede llegar a invadir el proceso de análisis de cesta de la compra si no se encuentra controlado el número de productos que se desean/pueden estudiar. A pesar de que diversos investigadores han desarrollado algoritmos para extraer asociaciones relevantes, existe otro método para abolir el problema de la dimensionalidad, el cual es la introducción de  $i$  productos dentro de un número  $j$  de categorías. Supongamos el caso de aquellos productos como por ejemplo el zumo que se encuentra en diversos tamaños en el punto de venta; para generalizar esta categoría en vez de denominarlos según su tamaño (bricks, bricks individuales...) o según su sabor (piña, melocotón, naranja...) se compactarán en una categoría general llamada zumos.

Al igual que ocurre en las organizaciones que ofrecen un elevado número de productos, la agregación de ítems también beneficia a aquellas cuyos datos, en principio, son reducidos, no siendo fiables las reglas de asociación. Por ejemplo, si solo existe una transacción con la marca A de zumo de naranja en una determinada cesta de la compra, la cual incluye también yogurt, la confianza que se obtendría de dicha asociación sería de 100 %; dicha asociación carece de interés. En este contexto, se destaca pues que cuantos más niveles de agregación se generan, el nivel de detalles aportados por el análisis disminuye, perdiendo información para el desarrollo de estrategias de marketing (Blattberg, Pyöng-do & Neslin, 2008).

## **8.3. HETEROGENEIDAD DE LOS CONSUMIDORES**

Un aspecto relevante que destacar en el presente escrito es la heterogeneidad que presentan los consumidores, es decir, las preferencias de los clientes son diferentes entre sí, y han de tenerse en cuenta. Las estrategias de marketing que se realicen serán pues más efectivas, debido a que el concepto de heterogeneidad es la base del proceso de segmentación, selección de público objetivo, posicionamiento de los productos y marcas y, también, del micro-marketing.

Se distinguen dos tipos de heterogeneidad: heterogeneidad de preferencia y heterogeneidad estructural. Por un lado, la heterogeneidad de preferencia hace referencia a las preferencias de cada consumidor en función de productos y marcas y las respuestas del consumidor en consecuencia de las estrategias implementadas de marketing-mix;

por otro lado, la heterogeneidad estructural estudia las diferencias estructurales existentes en el proceso de elección que realizan los clientes (Segarra Roca, 2007).

La heterogeneidad de preferencia destaca por su cobertura bibliográfica. En este contexto, algunos modelos manifiestan la heterogeneidad de preferencia de manera exógena, por ejemplo, con medias basadas en observaciones realizadas sobre decisiones de compra (Krishnamurthi & Raj, 1988) si bien determinados autores destacan a su vez las desventajas que presenta la heterogeneidad de preferencia que solo aprecia aspectos parciales y obvia los aspectos implícitos de los consumidores. Más aún, estableciendo la preferencia de marca como un indicador exógeno no se obtiene un análisis clarividente de las preferencias ni de las respuestas producidas en los consumidores a través de las variables de marketing-mix (Kamakura, Kim & Lee, 1996).

A diferencia que la heterogeneidad de preferencia, la heterogeneidad estructural se caracteriza por realizar un proceso de clasificación no paramétrica que culmina deduciendo las decisiones de los consumidores a través de una eliminación de manera jerárquica; se obtienen segmentos que recogen a consumidores con cestas de la compra similares. A lo largo de este proceso se tienen en cuenta no solo los niveles de preferencia de los clientes, sino además la estructura de los procesos de elección utilizados (Kannan & Wright, 1991).

En suma, la heterogeneidad debe ser concebida en el análisis cesta de la compra, puesto que permite desarrollar estrategias de marketing más adaptadas a los consumidores, añadiendo información sobre las preferencias de los consumidores. Véase este proceso como una mejora añadida a la información deducida a través del análisis cesta de la compra, estableciendo estrategias más eficaces en los puntos de ventas.

## **9. CASO PRÁCTICO 1: CESTA DE LA COMPRA EN SUPERMERCADOS**

El principal objetivo de este caso práctico reside en la obtención del conjunto de productos relevantes en un determinado supermercado, y utilizar la información extraída para la definición de perfiles de consumidores, aplicando mejoras en la distribución de los productos en el punto de venta.

En particular, el proceso de definir las transacciones que realizan los clientes es fundamental para definir el análisis de cesta de la compra en el área de marketing. En este sentido, una transacción típica es un conjunto de productos que ha comprado el consumidor de forma *online* u *offline*. La transacción genera datos que son introducidos en el software para posteriormente analizarlos. Los artículos que ha comprado un consumidor no solo se clasifican en función del producto que ha adquirido, sino también se estudia la información relativa a las organizaciones, industrias e incluso países. Cabe señalar que el periodo de recogida de transacciones puede ser un momento específico, un día entero o períodos largos de tiempo esta decisión será tomada en función del estudio que quiera realizar el investigador (Aguinis, Forcum & Joo, 2013).

A continuación, se desarrolla un caso práctico base que engloba un conjunto de datos perteneciente a un supermercado. La finalidad perseguida durante el análisis de dicha base de datos descansa en ofrecer una serie de recomendaciones que permitan al usuario entender el comportamiento del consumidor en puntos de ventas parecidos al que se despliega para así poder desarrollar una estrategia eficaz. No obstante, la importancia del caso práctico propuesto no reside en los conocimientos específicos para el

supermercado que se analiza, sino que el valor del presente caso descansa en las características de la metodología propuesta, siendo el enfoque utilizado aplicable a diversas industrias. Se pretende que la aplicación sea de valor para otros tipos de negocio o actividad comercial donde los consumidores adquieran productos o servicios.

Para realizar la parte práctica del actual escrito se ha utilizado la herramienta R, definida como un software de lenguaje de programación enfocado al análisis estadístico y gráfico. Su utilización resalta en el campo de la minería de datos, siendo RStudio el entorno de desarrollo del lenguaje de programación R. Ambos programas funcionan de manera conjunta. Asimismo, en el presente caso práctico se ha utilizado una base de datos secundaria de un supermercado real puesta a disposición de este tipo de estudios por el paquete R *arules* (Hashler, Gün & Hörnik, 2005), cuyo conjunto de datos contiene 30 días de transacciones en dicho punto de venta. La base de datos está compuesta por 9.835 transacciones y 169 categorías de productos (véase también, Hahsler, Hornik & Reutterer, 2006). Dentro de dicha base de datos se registra una serie de transacciones; cada transacción es identificada por un único número, y siendo cada transacción el conjunto de productos comprados por un cliente en un acto de compra concreto en el supermercado.

**Cuadro 2:** Extracto del conjunto de transacciones registradas.

Transacción ID	Productos
9.617	Carne de ternera, pan, licor, periódico, verduras, leche entera, yogurt
9.618	Café
9.619	bolsas, verduras encurtidas, carne de cerdo, bollería, golosinas, leche entera, yogurt

Fuente: Elaboración propia.

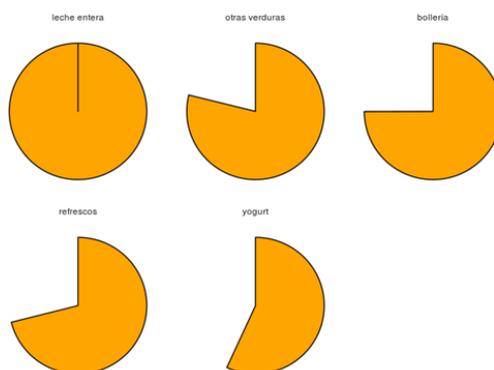
El Cuadro 2 es una ilustración de cómo se encuentran registradas las transacciones. Como se puede observar, cada transacción tiene asociado un número identificativo, al igual que se especifican los productos que el cliente ha adquirido en una compra puntual. Como se ha descrito anteriormente, se define soporte  $s$  como el número de transacciones para un determinado subconjunto de  $I$ . Por este motivo,  $I$  es frecuente si el soporte  $s$  es mayor que el denominado soporte mínimo. Asimismo, se debe determinar también el término confianza  $c$  (Videla-Cavieres & Rios, 2014). Aplicando el algoritmo *A priori* sobre la base de datos transaccionales que se están analizando, obtenemos de manera eficiente los conjuntos de categorías de productos más frecuentes de la base de datos.

Para poder aplicar dicho algoritmo se estima tanto el soporte como la confianza; el resultado es más preciso. En este caso se ha establecido un soporte = 0,005 y una confianza = 0,25, especificando que al menos debe haber dos productos en cada conjunto. Como resultado, el algoritmo *a priori* ha generado un total de 663 reglas, cada una de ellas con un mínimo de dos ítems.

La Figura 4 muestra las cinco categorías de productos que presentan mayor frecuencia relativa en la base de datos utilizada, deduciéndose que las categorías que aparecen son, en términos relativos, las categorías de productos que más se han adquirido; siendo la categoría leche entera la que obtiene el mayor valor de frecuencia relativa. En la

Tabla 5, se muestran las categorías más adquiridas durante la recogida de los datos, y sus valores absolutos.

**Figura 4:** Frecuencia relativa de las categorías de productos.



Fuente: Elaboración propia.

**Tabla 5:** Productos más frecuentes en la cesta de la compra de los encuestados.

Leche entera	Verduras	Bollería	Refresco	Yogurt
2.513	1.903	1.809	1.715	1.372

Fuente: Elaboración Propia.

Cabe señalar que en el presente método se han tenido que modificar en cada regla los niveles de soporte mínimo (*min\_supp*) y de confianza mínima (*min\_conf*) para poder encontrar las reglas que se muestran; los valores de dichos niveles se encuentran indicados.

### 9.1. LECHE ENTERA

Como se puede observar en la Tabla 6, los consumidores cuando compran fruta, hortalizas y ciertos lácteos tienden a comprar también leche entera. La mayor probabilidad de comprar leche sabiendo que se ha adquirido otra serie de productos se encuentra en el conjunto de productos formado por fruta tropical, hortalizas y verduras con un 70 % de probabilidad. Asimismo, la adquisición conjunta de leche entera, mantequilla y yogurt es igual al 0,94 % dentro del conjunto de transacciones.

**Tabla 6:** Tabla de asociaciones con leche entera en el lado derecho (RHS).

	LHS	RHS	SOPORTE	CONFIANZA	LIFT
1	{fruta tropical, hortalizas, yogurt}	{leche entera}	0,0057	0,70	2,7

2	{frutas con semillas, hortalizas, verduras}	{leche entera}	0,0055	0,68	2,6
3	{mantequilla, nata}	{leche entera}	0,0067	0,66	2,6
4	{frutas con semillas, nata}	{leche entera}	0,0060	0,65	2,5
5	{mantequilla, yogurt}	{leche entera}	0,0094	0,64	2,5

Fuente: Elaboración propia.

Min\_supp = 0,005, min\_conf = 0,05

La Tabla 7 muestra que los verduras son los productos con mayor probabilidad de ser adquiridos si (probabilidad condicional) se ha comprado leche entera. Esta asociación ha obtenido además un indicador lift de 1,513, determinándose pues una relación positiva, es decir, ambos productos son complementarios.

**Tabla 7:** Tabla de asociaciones con leche entera en el lado izquierdo (LHS).

	LHS	RHS	SOPORTE	CONFIANZA	LIFT
1	{leche entera}	{verduras}	0,074	0,292	1,513
2	{leche entera}	{bollería}	0,056	0,221	1,205
3	{leche entera}	{yogurt}	0,056	0,219	1,571
4	{leche entera}	{hortalizas}	0,048	0,191	1,756
5	{leche entera}	{fruta tropical}	0,042	0,165	1,577

Fuente: Elaboración propia.

Min\_supp = 0,001, min\_conf = 0,15

## 9.2. VERDURAS

Observando la Tabla 8, la asociación más significativa es la relación entre hortalizas y verduras; poseen una relación positiva. Del mismo modo, observando el indicador confianza se aprecia que cuando se compran hortalizas también se compran verduras, siendo catalogados como productos complementarios.

**Tabla 8:** Tabla de asociaciones con verduras en el lado derecho (RHS).

	LHS	RHS	SOPORTE	CONFIANZA	LIFT
1	{leche entera}	{verduras}	0,074	0,292	1,513
2	{hortalizas}	{verduras}	0,047	0,434	2,246
3	{yogurt}	{verduras}	0,043	0,311	1,608
4	{bollería}	{verduras}	0,042	0,231	1,197
5	{fruta tropical}	{verduras}	0,035	0,342	1,767

Fuente: Elaboración propia.

Min\_supp = 0,001, min\_conf = 0,10

Como muestra la Tabla 9, el yogurt tiene un comportamiento significativo en la asociación con las verduras; posee la mayor probabilidad de ser adquirido una vez se ha elegido el producto verduras; en su defecto, la categoría refresco tiene la menor probabilidad de ser elegido una vez se ha optado por las verduras. La asociación entre la categoría verduras y la categoría refresco presenta además un indicador lift menor que 1.

**Tabla 9:** Tabla de asociaciones con verduras en el lado izquierdo (LHS).

	LHS	RHS	SOPORTE	CONFIANZA	LIFT
1	{verduras}	{yogurt}	0,0434	0,224	1,608
2	{verduras}	{bollería}	0,042	0,220	1,197
3	{verduras}	{fruta tropical}	0,035	0,185	1,767
4	{verduras}	{refresco}	0,032	0,169	0,970

Fuente: Elaboración propia.

Min\_supp = 0,005, min\_conf = 0,15

### 9.3. BOLLERÍA

La Tabla 10 muestra como el conjunto de categoría de productos compuesto por margarina, queso y periódicos es la cesta con mayor probabilidad de adquirir bollería.

**Tabla 10:** Tabla de asociaciones con bollería en el lado derecho (RHS).

	LHS	RHS	SOPORTE	CONFIANZA	LIFT
1	{margarina, queso, periódicos}	{bollería}	0,0012	0,75	4,077
2	{carne de ternera, fruta tropical, leche entera, yogurt}	{bollería}	0,0013	0,684	3,719
3	{cítricos, leche entera, nata, hojaldre}	{bollería}	0,0010	0,667	3,624
4	{refresco, chocolate, golosinas}	{bollería}	0,0012	0,6316	3,433
5	{verduras, chocolate, servilletas}	{bollería}	0,0010	0,625	3,397

Fuente: Elaboración propia.

Min\_supp=0,001, min\_conf = 0,01

Como refleja la Tabla 11, si se adquiere la categoría bollería también se adquiere la categoría leche entera. Asimismo, el indicador soporte es el que tiene mayor probabilidad en este caso; ambos productos se consideran productos complementarios, puesto que el indicador lift muestra una relación de interdependencia mayor que 1.

Las categorías de productos bollería y salchichas en la Tabla 11 mantienen una relación de interdependencia, si bien el indicador confianza muestra que la probabilidad de ser

adquirida la categoría salchichas una vez ha sido seleccionada la categoría bollería es baja.

**Tabla 11:** Tabla de asociaciones con bollería en el lado izquierdo (LHS).

	LHS	RHS	SOPORTE	CONFIANZA	LIFT
1	{bollería}	{leche entera}	0,056	0,308	1,205
2	{bollería}	{verdura}	0,043	0,232	1,197
3	{bollería}	{refresco}	0,038	0,208	1,195
4	{bollería}	{yogurt}	0,034	0,187	1,339
5	{bollería}	{salchichas}	0,031	0,166	1,771

Fuente: Elaboración propia.

Min\_supp=0,001, min\_conf = 0,15

#### 9.4. REFRESCO

La Tabla 12 muestra que la probabilidad de escoger la categoría de producto refresco una vez han sido seleccionada las categorías de productos café y alcohol es elevada, 76,9 %. Se observa también que el indicador lift es igual a 4,411, siendo categorías complementarias.

**Tabla 12:** Tabla de asociaciones con refresco en el lado derecho (RHS).

	LHS	RHS	SOPORTE	CONFIANZA	LIFT
1	{café, alcohol}	{refresco}	0,001	0,769	4,411
2	{yogurt, bollería, agua, periódicos}	{refresco}	0,001	0,769	4,411
3	{salchichas, agua, cerveza}	{refresco}	0,001	0,734	4,205
4	{salchichas, pan, bolsa de la compra}	{refresco}	0,001	0,667	3,823
5	{bollería, agua, chocolate}	{refresco}	0,001	0,650	3,727

Fuente: Elaboración propia.

Min\_supp=0,001, min\_conf = 0,08

Observando la Tabla 13, se obtiene que la relación entre la categoría refresco y la categoría leche entera posee un lift inferior a 1, por consiguiente, el nivel de interdependencia es bajo y se tildan como productos sustitutos. Sin embargo, la probabilidad de adquirir la categoría leche entera cuando ha sido seleccionada la categoría refresco obtiene el porcentaje más alto de la Tabla 13.

Asimismo, se observa que las categorías de productos refresco y agua son consideradas categorías complementarias; el valor de lift es mayor que 1.

**Tabla 13:** Tabla de asociaciones con refresco en el lado izquierdo (LHS).

	LHS	RHS	SOPORTE	CONFIANZA	LIFT
1	{refresco}	{leche entera}	0,040	0,229	0,899
2	{refresco}	{bollería}	0,0383	0,219	1,195
3	{refresco}	{verduras}	0,0327	0,187	0,970
4	{refresco}	{agua}	0,0289	0,166	1,503
5	{refresco}	{yogurt}	0,027	0,157	1,124

Fuente: Elaboración propia.

Min\_supp=0,001, min\_conf = 0,1

## 9.5. YOGURT

Cuando se compra fruta tropical existe una alta probabilidad (27,907 %) de seleccionar el producto yogurt; además ambos productos poseen una relación de interdependencia muy positiva, como se refleja en la tabla 19.

**Tabla 14:** Tabla de asociaciones con yogurt en el lado derecho (RHS).

	LHS	RHS	SOPORTE	CONFIANZA	LIFT
1	{leche entera}	{yogurt}	0,056	0,219	1,572
2	{verduras}	{yogurt}	0,043	0,224	1,608
3	{bollería}	{yogurt}	0,034	0,187	1,340
4	{fruta tropical}	{yogurt}	0,029	0,279	2,0

Fuente: Elaboración propia.

Min\_supp=0,005, min\_conf = 0,05

La Tabla 15 resalta que cuando se adquiere el producto yogurt se compra conjuntamente el producto leche entera. Los tres indicadores (soporte, confianza y lift) muestran buenos resultados en dicha combinación. Se explica este hecho en que siendo lácteos, pero no comparten necesariamente la misma finalidad para el consumidor.

**Tabla 15:** Tabla de asociaciones con yogurt en el lado izquierdo (LHS).

	LHS	RHS	SOPORTE	CONFIANZA	LIFT
1	{yogurt}	{leche entera}	0,056	0,402	1,570
2	{yogurt}	{verduras}	0,043	0,311	1,608
3	{yogurt}	{bollería}	0,034	0,246	1,339
4	{yogurt}	{fruta tropical}	0,029	0,209	2,0

5	{yogurt}	{refresco}	0,027	0,196	1,124
---	----------	------------	-------	-------	-------

Fuente: Elaboración propia.

Min\_supp=0,001, min\_conf = 0,15

Identificados y analizados los productos que más se han adquirido en este conjunto de transacciones de cesta de la compra, se continúa analizando una serie de categorías que se han considerado relevantes para completar el presente manuscrito.

## 9.6. ARTÍCULOS DE HIGIENE

Como se observa en la Tabla 16 los artículos de higiene poseen una asociación muy positiva de interdependencia con categorías *alejadas* (véase Tabla 16). Este hecho evidencia que los artículos de higiene sean considerados productos de primera necesidad. Por ejemplo, cuando se ha elegido leche entera, mantequilla y servilletas la probabilidad de adquirir artículos higiene es igual al 32,2 %.

Como se refleja en la Tabla 17, los artículos de higiene y leche entera obtienen un valor lift mayor que 1, traduciéndose este hecho en que poseen una interdependencia positiva. Véase que además el indicador confianza adquiere el mayor valor de la Tabla 17 para dicha asociación, siendo la probabilidad de adquirir la categoría leche entera una vez ha sido seleccionada la categoría artículos de higiene la más elevada.

**Tabla 16:** Tabla de asociaciones con artículos de higiene en el lado derecho (RHS).

	LHS	RHS	SOPORTE	CONFIANZA	LIFT
1	{leche entera, mantequilla, servilletas}	{artículos de higiene}	0,001	0,322	9,792
2	{leche entera, agua, servilletas}	{artículos de higiene}	0,001	0,282	8,562
3	{cuajada, servilletas}	{artículos de higiene}	0,001	0,277	8,396
4	{fruta tropical, leche entera, servilletas}	{artículos de higiene}	0,001	0,245	7,434
5	{salchicha, fruta con semillas, leche entera}	{artículos de higiene}	0,001	0,236	7,175

Fuente: Elaboración propia.

Min\_supp=0,001, min\_conf = 0,08

**Tabla 17:** Tabla de asociaciones con artículos de higiene en el lado izquierdo (LHS).

	LHS	RHS	SOPORTE	CONFIANZA	LIFT
1	{artículos de higiene}	{leche entera}	0,013	0,3889	2,0
2	{artículos de higiene}	{verduras}	0,009	0,290	1,499

3	{artículos de higiene}	{yogurt}	0,007	0,222	1,593
4	{artículos de higiene}	{refresco}	0,007	0,213	1,221
5	{artículos de higiene}	{fruta tropical}	0,006	0,204	1,941

Fuente: Elaboración propia.

Min\_supp=0,001, min\_conf = 0,15

## 9.7. ARROZ

La siguiente categoría de producto que ha sido analizada es el arroz, debido a que es considerado un cereal básico en la dieta. Este hecho se observa en el valor lift en cada serie de productos que forman la Tabla 18. En particular, se destaca el conjunto de categorías formado por hortalizas, verduras y yogurt. Se traduce en que la relación de dichos productos guarda una relación de interdependencia con el arroz, demostrando ser productos complementarios entre sí.

Asimismo, la relación se describe también mediante el porcentaje del indicador confianza; existe un 11,023 % de probabilidad de adquirir arroz una vez se han elegido hortalizas, verduras y yogurt.

**Tabla 18:** Tabla de asociaciones con arroz en el lado derecho (RHS).

	LHS	RHS	SOPORTE	CONFIANZA	LIFT
1	{hortalizas, verduras, leche entera}	{arroz}	0,002	0,078	10,353
2	{hortalizas, yogurt}	{arroz}	0,002	0,063	8,260
3	{leche entera, mantequilla}	{arroz}	0,002	0,055	7,258
4	{verduras, leche entera, yogurt}	{arroz}	0,002	0,068	8,982
5	{hortalizas, verduras, yogurt}	{arroz}	0,001	0,110	14,456

Fuente: Elaboración propia.

Min\_supp=0,001, min\_conf = 0,05

Para analizar la Tabla 19, es decir, donde el arroz ocupa el lado izquierdo de la tabla, se observa el indicador confianza; el mayor porcentaje de este indicador se encuentra en la relación entre el producto arroz y leche entera. Adicionalmente, se denota una alta probabilidad de este indicador en todas las asociaciones reflejadas en la tabla. Basándose en el indicador soporte, se puede argumentar que la asociación que conforma el arroz y la leche entera poseen, nuevamente, el porcentaje más alto en este análisis, desembocando este porcentaje en la explicación de que existe una alta probabilidad, en comparación con el resto de las asociaciones analizadas, de adquirir arroz y leche entera de manera simultánea en una misma cesta de la compra.

**Tabla 19:** Tabla de asociaciones con arroz en el lado izquierdo (LHS).

	LHS	RHS	SOPORTE	CONFIANZA	LIFT
1	{arroz}	{leche entera}	0,005	0,613	2,40
2	{arroz}	{verduras}	0,004	0,520	2,687
3	{arroz}	{hortalizas}	0,003	0,413	3,792
4	{arroz}	{yogurt}	0,002	0,307	2,198
5	{arroz}	{zumos}	0,002	0,253	3,504

Fuente: Elaboración propia.

Min\_supp=0,001,min\_conf = 0,10

### 9.8. TERNERA

Analizando la Tabla 20 se observa que el mayor porcentaje del indicador confianza reside en el conjunto formado por la fruta tropical, hortalizas, verduras, leche entera y bollería. Pone de manifiesto que una vez seleccionado dicho conjunto de productos existe un 55 % de probabilidad de adquirir también el producto ternera. De la misma manera, se posee el valor lift superior.

Cuando la categoría de producto ternera se encuentra en el lado izquierdo (LHS) de la Tabla 21, los indicadores muestran variaciones relevantes, debido a que el producto ternera es seleccionado al principio y el resto de los productos es influido por dicha decisión. La asociación que presenta la ternera y la categoría de producto hortaliza destaca por poseer el mayor indicador lift, es decir, mayor interdependencia, evidenciando dicho nivel que ambas categorías sean consideradas como complementarias.

Asimismo, se observa que la asociación de ternera y leche entera es más probable de ser adquirida conjuntamente; igualmente, esta asociación refleja la mayor confianza - una vez se ha escogida la categoría ternera existe un 40,5 % de probabilidades de adquirir la categoría de producto leche entera.

**Tabla 20:** Tabla de asociaciones con ternera en el lado derecho (RHS).

	LHS	RHS	SOPORTE	CONFIANZA	LIFT
1	{fruta tropical, hortalizas, verduras, leche entera, bollería}	{ternera }	0,001	0,550	10,483
2	{hortalizas, leche entera, mantequilla, bollería}	{ternera }	0,001	0,478	9,116
3	{salchicha, hortalizas, mantequillas}	{ternera }	0,001	0,455	8,664
4	{hortalizas, mantequilla, yogurt}	{ternera }	0,002	0,395	7,524
5	{fruta tropical, hortalizas, leche entera, bollería}	{ternera }	0,001	0,378	7,212

Fuente: Elaboración propia.

Min\_supp=0,001, min\_conf = 0,08

**Tabla 21:** Tabla de asociaciones con ternera en el lado izquierdo (LHS).

	LHS	RHS	SOPORTE	CONFIANZA	LIFT
1	{ternera}	{leche entera}	0,021	0,405	1,585
2	{ternera}	{verduras}	0,019	0,376	1,943
3	{ternera}	{hortalizas}	0,017	0,331	3,040
4	{ternera}	{bollería}	0,014	0,259	1,412
5	{ternera}	{yogurt}	0,012	0,223	1,598

Fuente: Elaboración propia.

Min\_supp=0,001,min\_conf = 0,15

## 9.9. HUEVOS

La asociación con mayor porcentaje en el indicador soporte es el conjunto leche entera y huevos (2,9 %). La asociación formada por hortalizas y huevos obtiene los mayores resultados en los indicadores lift y confianza, categorías de productos que presentan dependencia en la cesta de la compra; estas categorías son consideradas complementarias entre sí.

Referenciando la Tabla 23 se observa que los huevos junto con la categoría de producto leche entera muestran la probabilidad más alta tanto de ser adquiridos de forma conjunta, y de ser elegida la leche entera una vez seleccionados los huevos.

Se vuelve a poner de manifiesto en la Tabla 23, la interdependencia que muestran las categorías de productos huevos y hortalizas, siendo pues calificadas como categorías complementarias.

**Tabla 22:** Tabla de asociaciones con huevos en el lado derecho (RHS).

	LHS	RHS	SOPORTE	CONFIANZA	LIFT
1	{leche entera}	{huevos }	0,029	0,117	1,850
2	{verduras}	{huevos }	0,022	0,115	1,814
3	{bollería}	{huevos }	0,016	0,085	1,342
4	{hortalizas}	{huevos }	0,014	0,132	2,073
5	{yogurt}	{huevos }	0,014	0,103	1,619

Fuente: Elaboración propia.

Min\_supp=0,008, min\_conf = 0,08

**Tabla 23:** Tabla de asociaciones con huevos en el lado izquierdo (LHS).

	LHS	RHS	SOPORTE	CONFIANZA	LIFT
1	{huevos}	{leche entera}	0,029	0,473	1,850

2	{huevos}	{verduras}	0,0221	0,351	1,814
3	{huevos}	{bollería}	0,016	0,247	1,342
4	{huevos}	{hortalizas}	0,014	0,226	2,073
5	{huevos}	{yogurt}	0,014	0,226	1,619

Fuente: Elaboración propia.

Min\_supp=0,001, min\_conf = 0,09

### 9.10. PESCADO ENVASADO

En la Tabla 24 se observa la relación positiva entre la leche entera y los artículos de higiene con el pescado envasado.

**Tabla 24:** Tabla de asociaciones con pescado envasado en el lado derecho (RHS).

	LHS	RHS	SO- PORTE	CONFIANZA	LIFT
1	{leche entera, artículos de higiene}	{pescado envasado}	0,001	0,087	5,801
2	{mostaza}	{pescado envasado}	0,001	0,085	5,632
3	{verduras, café}	{pescado envasado}	0,001	0,076	5,034
4	{verduras, verduras congelados}	{pescado envasado}	0,001	0,0628	4,177
5	{encurtidos}	{pescado envasado}	0,001	0,0625	4,153

Fuente: Elaboración propia.

Min\_supp=0,001, min\_conf = 0,01

Como se refleja en la Tabla 25, una vez ha sido seleccionado el producto pescado envasado existe una probabilidad del 33,7 % de seleccionar verduras. Es la asociación entre las categorías de producto pescado envasado y pan integral las que obtienen la relación más alta y superior a 1, un valor igual a 2,396.

**Tabla 25:** Tabla de asociaciones con pescado envasado en el lado izquierdo (LHS).

	LHS	RHS	SOPORTE	CONFIANZA	LIFT
1	{pescado envasado}	{pan integral}	0,002	0,155	2,396
2	{pescado envasado}	{bolsa de la compra}	0,003	0,196	1,989
3	{pescado envasado}	{verduras}	0,005	0,3388	1,746
4	{pescado envasado}	{hortalizas}	0,003	0,176	1,612

5	{pescado envasado}	{yogurt}	0,003	0,216	1,549
---	--------------------	----------	-------	-------	-------

Fuente: Elaboración propia.

Min\_supp=0,002, min\_conf = 0,15

### 9.11. PAN INTEGRAL

Haciendo referencia a la Tabla 26, se puede concluir que cuando la categoría de producto pan integral ocupa el lado derecho (RHS) de la tabla, la probabilidad de escoger dicha categoría de producto una vez seleccionado el conjunto salchicha y verduras es igual a 16,6 %. Las categorías de leche entera y huevos evidencian la probabilidad más baja de elegir seguidamente pan integral. Adicionalmente, el indicador soporte resalta la asociación que existe entre leche entera y hojaldre junto al pan integral, indicando que la probabilidad de comprarlos simultáneamente es la más elevada.

**Tabla 26:** Tabla de asociaciones con pan integral en el lado derecho (RHS).

	LHS	RHS	SOPORTE	CONFIANZA	LIFT
1	{salchicha, verduras }	{pan integral}	0,004	0,166	2,559
2	{salchicha, leche entera}	{pan integral}	0,004	0,149	2,307
3	{leche entera, periódico}	{pan integral}	0,004	0,148	2,292
4	{leche entera, hojaldre}	{pan integral}	0,005	0,143	2,215
5	{leche entera, huevos}	{pan integral}	0,004	0,135	2,090

Fuente: Elaboración propia.

Min\_supp=0,004, min\_conf = 0,04

Cuando la categoría de producto pan integral ocupa el lado izquierdo de la tabla, véase la Tabla 27, se observa que la relación existente entre las categorías de productos pan integral y leche entera obtiene los valores más altos de los tres indicadores expuestos. En su defecto, la asociación que forman las categorías de productos pan integral y bollería han adquirido los valores más bajos.

**Tabla 27:** Tabla de asociaciones con pan integral en el lado izquierdo (LHS).

	LHS	RHS	SOPORTE	CONFIANZA	LIFT
1	{pan integral}	{leche entera}	0,025	0,388	1,521
2	{pan integral}	{verduras}	0,019	0,288	1,490
3	{pan integral}	{yogurt}	0,014	0,224	1,606
4	{pan integral}	{refresco}	0,012	0,194	1,114
5	{pan integral}	{bollería}	0,012	0,194	1,056

Fuente: Elaboración propia.

Min\_supp=0,001, min\_conf = 0,15

## 9.12. ASOCIACIONES SEGÚN LIFT

La tabla 28 recoge las 10 asociaciones con lift más significativo del conjunto de transacciones. Esta manera de analizar el conjunto de reglas de asociación resulta interesante para el presente análisis; muestra cuáles son los productos que poseen mayor conexión entre ellos. La finalidad perseguida reside en la adquisición del conocimiento general sobre la cesta de la compra que realiza el consumidor.

Para ejemplificar lo expuesto anteriormente, la asociación con lift más elevado en la base de datos que se está analizando se encuentra en la regla que asocia el conjunto de categorías de productos cítricos, verduras y leche entera con la categoría hortalizas. El cliente que adquiere cítricos, verduras y leche entera evidencia un lift superior a 4.

**Tabla 28:** Tabla de asociaciones con lift más significativos.

	LHS	RHS	SOPORTE	CONFIANZA	LIFT
1	{cítricos, verduras, leche entera}	{hortalizas}	0,005	0,445	4,085
2	{mantequilla, verduras }	{nata}	0,005	0,289	4,036
3	{especias}	{hortalizas}	0,007	0,431	3,956
4	{cítricos, frutas con semillas}	{fruta tropical}	0,005	0,404	3,854
5	{frutas del bosque}	{nata}	0,009	0,272	3,796
6	{ verduras, fruta tropical, leche entera}	{hortalizas}	0,007	0,410	3,768
7	{hortalizas, leche entera, yogurt}	{fruta tropical}	0,005	0,391	3,732
8	{ verduras, frutas con semillas, leche entera}	{hortalizas}	0,005	0,406	3,724
9	{cítricos, fruta tropical}	{frutas con semillas}	0,005	0,280	3,709
10	{cuajada, fruta tropical}	{yogurt}	0,005	0,5145	3,690

Fuente: Elaboración propia.

## 9.13. ASOCIACIONES SEGÚN SOPORTE

El indicador soporte, como se ha mostrado previamente en este escrito, manifiesta la probabilidad de adquirir una serie de productos de forma conjunta. Haciendo referencia a la Tabla 29, se observa como las categorías de productos leche entera y verduras ocupan los dos primeros puestos, siendo la probabilidad de ser adquiridos de manera simultánea igual a 7.4 %.

**Tabla 29:** Tabla de asociaciones con soporte más significativos.

	LHS	RHS	SOPORTE	CONFIANZA	LIFT
1	{verduras}	{leche entera}	0,074	0,386	1,513
2	{leche entera}	{verduras}	0,074	0,292	1,513
3	{bollería}	{leche entera}	0,057	0,307	1,205
4	{yogurt}	{leche entera}	0,056	0,401	1,571
5	{hortalizas}	{leche entera}	0,049	0,448	1,756
6	{hortalizas}	{verduras}	0,047	0,434	2,246
7	{yogurt}	{verduras}	0,043	0,31	1,608
8	{fruta tropical}	{leche entera}	0,042	0,403	1,577
9	{fruta tropical}	{verduras}	0,035	0,342	1,767
10	{agua embotellada}	{leche entera}	0,034	0,310	1,216

Fuente: Elaboración propia.

#### 9.14. ASOCIACIONES SEGÚN CONFIANZA

No obstante, obtener la Tabla 30 donde se refleja el conjunto de asociaciones con valores más significativos en el indicador confianza aporta interés al análisis. El indicador confianza de una asociación muestra el número de transacciones que contienen tanto antecedentes como consecuentes por lo que se considera una medida de precisión (Rompré, Biskri & Meunier, 2017). Como refleja la Tabla 30, en el presente caso, las categorías de productos hortalizas, fruta tropical y yogurt presentan la probabilidad más destacada, es decir, 70 % de probabilidades de adquirirse conjuntamente con la leche entera.

**Tabla 30:** Tabla de asociaciones con confianza más significativas.

	LHS	RHS	SOPORTE	CONFIANZA	LIFT
1	{hortalizas, fruta tropical, yogurt}	{leche entera}	0,006	0,70	2,73
2	{verduras, fruta con semillas, hortalizas}	{leche entera}	0,005	0,6750	2,64
3	{mantequilla, nata}	{leche entera}	0,007	0,660	2,58
4	{fruta con semillas, nata }	{leche entera}	0,006	0,648	2,537
5	{mantequilla, yogurt}	{leche entera}	0,009	0,638	2,50
6	{mantequilla, hortalizas}	{leche entera}	0,008	0,637	2,496
7	{cuajada, fruta tropical}	{leche entera}	0,006	0,633	2,479

8	{cítricos, hortalizas, leche entera}	{verduras}	0,005	0,633	3,273
9	{verduras, fruta con semilla, yogurt}	{leche entera}	0,005	0,6250	2,446
10	{huevos, fruta con semillas}	{leche entera}	0,005	0,623	2,440

Fuente: Elaboración propia.

### 9.15. TÉCNICAS DE MINERÍA GRÁFICA

Durante este escrito se ha resaltado la finalidad que concierne al análisis cesta de la compra, y la existencia de diversas técnicas para conseguir descubrir los conjuntos de ítems que se adquieren de forma frecuente en los puntos de ventas. Por consiguiente, en el presente apartado, se propone una metodología basada en detectar dichos conjuntos de productos, tildados de frecuentes, a través de la representación gráfica de redes de productos y comunidades superpuestas.

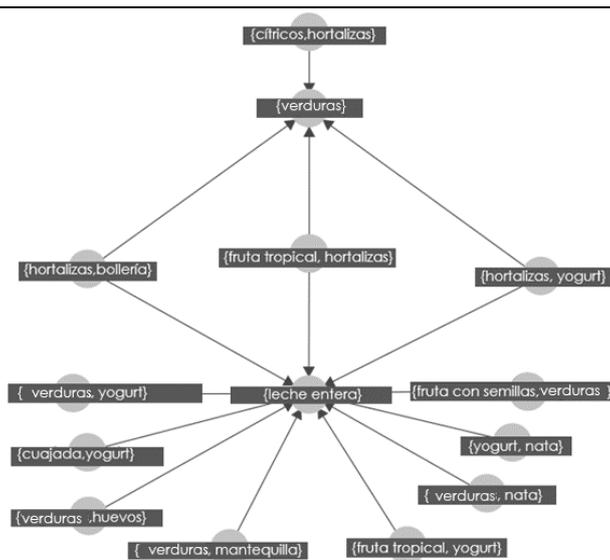
Sea una red definida como una serie de ítems interconectados entre ellos; siendo una manera común de representación de dicha red el uso de la gráfica. En consecuencia, una gráfica es una manera de especificar las relaciones que se presentan entre los conjuntos de productos; una gráfica está compuesta por una serie de productos que son llamados *nódulos*, conectados entre ellos, si así se requiere, por vínculos conocidos como *bordes*. Se deduce, por tanto, que una red de productos es una red donde los denominados *nódulos* representan productos mientras que los *bordes* representan las relaciones que existen entre diversos de *nódulos* (Videla-Cavieres & Rios, 2014).

Existen diversas formas para construir redes de productos. En primer lugar, se puede plantear crear una red de productos bipartitas caracterizada por unir productos y consumidores, intentando conseguir visualizar una relación conjunta entre todos los productos (Kim, Kim & Chen, 2012). Sin embargo, este método parece estar limitado por la necesidad de identificar a cada consumidor con su compra correspondiente, perdiendo información valiosa del conjunto de transacciones que se encuentra en la base de datos. En segundo lugar, como exponen Raeder & Chawla (2009), los datos utilizados para construir la red de productos estarían basada únicamente en las transacciones; siendo cada producto relacionado con otro producto. Consecuentemente, se buscan comunidades de productos que posean una fuerte relación entre sí, para dar soporte visual a las asociaciones que presentan una serie de productos.

### 9.16. TÉCNICAS DE MINERÍA GRÁFICA APLICADAS AL CASO PRÁCTICO 1

Una manera de apreciar las asociaciones que han sido obtenidas en el análisis cesta de la compra es a través de las técnicas de minería gráfica. La Figura 4 representa una red, es decir, un conjunto de categorías de productos interconectados, donde se especifica las relaciones entre las mismas. Cada nodo representa una categoría de producto perteneciente a la base de datos; siendo la conexión entre dichas categorías las flechas que aparecen en la Figura 4. En este contexto, las flechas dirigidas indican la regla de asociación que ha sido formada, coincidiendo las flechas que apuntan desde los nodos hasta los bordes con el lado derecho (RHS) de la asociación desembocando dicha flecha en el lado izquierdo (LHS).

**Figura 5:** Asociaciones a través de minería gráfica.



Fuente: Elaboración propia.

### 9.17. CONCLUSIÓN CASO PRÁCTICO 1

Durante este caso práctico, se han corroborado asociaciones entre categorías de productos adquiridos en un supermercado. Este hecho fundamenta la creciente importancia que recae en el análisis de la gestión, debido al interés en comprender el fenómeno que causa dichas asociaciones. De acuerdo con el análisis de cesta de la compra, las propuestas explicativas pueden ser desarrolladas no solo con datos transversales sino longitudinales.

Dentro del desarrollo de las propuestas se pueden diferenciar dos tipos de análisis de cesta de la compra: análisis cesta de la compra múltiple o análisis cesta de la compra secuencial. Por un lado, el análisis cesta de la compra múltiple hace referencia a que los datos disponibles incluyen transacciones que se han repetido a lo largo del tiempo. La primera aproximación se realiza con un estudio de los datos, concluyendo con el examen de los tres indicadores para determinar si se han producido cambios significativos en sus valores durante el periodo analizado. Por otro lado, el análisis de cesta de la compra secuencial es usado cuando los datos disponibles permiten describir eventos puntuales que se han repetido en el tiempo (Aguinis, Forcum & Joo, 2013).

El análisis cesta de la compra que ha sido desarrollado en el presente manuscrito se categoriza como un análisis descriptivo, puesto que la finalidad perseguida durante el estudio de las asociaciones ha sido esbozar las propiedades que se manifiestan en la cesta de la compra de la población analizada, permitiendo pues conocer las categorías de producto que se adquieren en un supermercado. Como ejemplo ilustrativo, se pueden apreciar diversas asociaciones asiduas. Véase que el mayor porcentaje obtenido para el indicador confianza corresponde a la asociación entre {café, alcohol} y {refresco}. Esta asociación se ha obtenido a través de un elevado volumen de registros de compras de consumidores, permitiendo al supermercado desarrollar una estrategia en la gestión del punto de venta, por ejemplo, realizar una promoción de descuentos en uno de los productos. Como se ha mencionado anteriormente, son complementarios

y dicha promoción acrecentará la adquisición de ambos productos. También, cabe distribuir ambos productos en diferentes puntos del establecimiento para que el consumidor lo recorra.

El estudio mostrado en este primer caso práctico incita pues a aplicar técnicas de merchandising que proporcionen una manera eficaz de gestionar el punto de venta. En este sentido, la disposición interna del punto de venta es considerada un elemento tanto informativo como orientativo para los consumidores, y facilita la compra de los clientes aumentando pues el nivel de satisfacción de los mismos. Más aún, una de las materias que concierne al estudio del merchandising es el itinerario o ruta seguido por el consumidor en el establecimiento durante su proceso de compra. El diseño del itinerario se refuerza por el análisis cesta de compra que permite conocer los bienes que adquieren los clientes y, por tanto, esbozar el recorrido que pueden proponer en el establecimiento para maximizar la utilidad.

En este contexto, cabe diferenciar las compras dirigidas por un objetivo y las compras por placer o exploratorias. En primer lugar, las compras dirigidas por un objetivo se caracterizan por ser compras de conveniencia. El cliente valora el tiempo invertido en realizarla; se busca que las compras sean eficientes. En segundo lugar, las compras como placer se asocian al *ambiente* generado en el establecimiento y el disfrute intrínseco evocado durante el ejercicio de la actividad. A diferencia del anterior tipo de compra, el factor tiempo es menos relevante (Díez, Landa & Navarro, 1996).

Asimismo, habiendo definido el tipo de compra que realizan los consumidores en el establecimiento, cabe señalar que posee ciertas zonas preferentes. Es lo que se conoce en la arquitectura del establecimiento comercial como puntos calientes; del mismo modo se encuentran los denominados puntos fríos. Los mencionados puntos calientes son espacios que propician flujos y concentración de clientes; por ejemplo, los puntos de acceso al mismo. Los puntos fríos se corresponden a las zonas menos accesibles y menos visuales de la superficie de ventas; por ejemplo, los ángulos poligonales del área perimetral del establecimiento (Borja, 2009). Cabe resaltar que los consumidores tienden a recorrer las zonas calientes del punto de venta, por lo que se requiere que existan el mayor número de zonas calientes en el establecimiento.

Las zonas calientes pueden diferenciarse en zonas calientes naturales, aquellas pertenecientes a la arquitectura del establecimiento, o zonas calientes artificiales, aquellas zonas que necesitan aplicar técnicas de merchandising (Martínez, 2016). Véase que en supermercados como Mercadona o Supersol, entre otros, la localización de los productos lácteos se encuentra en localizaciones alejadas de la zona de entrada; si bien promueven un proceso tedioso de búsqueda, permiten 'calentar' zonas frías naturales. Reside la explicación de este suceso en que la categoría leche entera, como se ha observado durante el análisis previo, es el producto más adquirido. Los consumidores transitarán el recorrido establecido para adquirir esta categoría de producto.

Aplicando estos comentarios al caso práctico 1, se recomienda poner los productos que son considerados productos de compra habitual, e.g. leche entera, en puntos fríos del establecimiento comercial; el consumidor hará un esfuerzo por adquirir dicho producto, y así se incrementará el flujo de circulación.

Más aún se deben categorizar los productos en función de si son necesarios o deseados y, además, si la compra se realiza de forma planificada. Este tipo de productos crea un recorrido en el establecimiento cuyo desplazamiento permite observar otros tipos de

productos. Se colige que este tipo de productos debe localizarse en los denominados puntos fríos del área comercial. Por otro lado, las denominadas compras de productos necesarios no planificadas corresponden a categorías de productos necesarios que el cliente adquiere en el punto de venta sin planificación previa; para que se realice la compra por el producto debe situarse en una posición visible y, además, en zona próxima a productos necesarios de compra planificada. Haciendo alusión al caso práctico 1 se puede observar que la categoría de producto arroz una vez ha sido adquirido muestra una probabilidad del 52 % de que también se adquirieran verduras; se recomendaría que ambas categorías estuvieran colocadas de forma próxima en la denominada zona fría, son productos categorizados de primera necesidad, siendo la categoría arroz una compra planificada, mientras que la categoría verduras, a pesar de ser una categoría de primera necesidad, se tilda como una compra no planificada condicionada por la adquisición de la categoría arroz.

Finalmente, las acciones promocionales se representan como estímulos de carácter psicológico que influyen en las dimensiones cognitiva, dimensión afectiva y comportamental. En este sentido, la dimensión cognitiva hace referencia a la captación de la atención del consumidor; la forma de presentación del producto es esencial. Por otro lado, la dimensión afectiva busca persuadir al consumidor, añadiendo un valor determinado a la oferta; se asume que dicho plus es el centro de la estrategia promocional. Por último, la dimensión comportamental se caracteriza por afectar la conducta del consumidor, utilizando la comunicación de la promoción de incentivo para cumplir con el objetivo propuesto (Díez, Landa & Navarro, 1996). Se deduce que las respuestas de los consumidores estarán determinadas por comportamientos de aproximación o, por el contrario, de rechazo hacia las promociones, siendo la consecuencia de dichos comportamientos los estados afectivos y cognitivos de los consumidores (Park & Lennon, 2009). Del mismo modo, la intención de compra puede estar influida por la disposición de productos en el punto de venta, puesto que añade valor percibido, incrementando la predisposición de compra (Bloemer & De Ruyter, 1998).

Existen diversos tipos de promociones, sin embargo, en el presente escrito se va a profundizar en el tipo de promoción conocido como venta conjunta, es decir, la venta de dos o más productos que son considerados unidades independientes, siendo el precio de dichos productos inferior si se compara con el precio de venta por separados. Existen diversas maneras de presentar este tipo de promoción, por ejemplo, vender por lotes los productos considerados de gran consumo, por ejemplo, packs de Coca-Cola; asociar artículos posicionados en la mente del consumidor con otros no muy conocidos; o añadir un artículo promocional al producto. Se resalta la importancia de no promocionar productos muy parecidos o que satisfagan las mismas necesidades, podrían ocasionar la anulación de los efectos perseguidos cuando se aplica la técnica de promoción (Wellhoff & Masson, 2005).

Véase que el caso práctico 1 permite diseñar promociones para incentivar las ventas de determinadas asociaciones de productos. Los siguientes ejemplos que se proponen se basan en el índice de confianza, su definición hace alusión a la elección de un producto una vez ha sido seleccionado previamente un producto o un conjunto de productos.

En primer lugar, si se analiza la categoría de producto leche entera cuando han sido seleccionadas las categorías de productos mantequilla y yogurt, la probabilidad de adquirir la categoría leche entera es del 64 %; aplicar una promoción podría disminuir la

percepción de valor por parte del consumidor, debido a que la reducción del precio puede ser percibida por el consumidor como una reducción del valor del producto. Sin embargo, cabe señalar que si una marca específica de leche entera observase que las ventas están disminuyendo podría plantearse crear una promoción específica con una marca de mantequilla o yogurt para aumentar la adquisición por parte de los consumidores de su marca; este tipo de promoción recibe el nombre de *prima autofinanciada*. Se precisa que los productos pueden no pertenecer a la misma categoría; véase el caso en que una vez ha sido seleccionada la categoría de producto leche entera existe una probabilidad de 16,55 % de adquirir también fruta tropical, dicha probabilidad puede ser incrementada si se aplica una promoción.

## **10. SISTEMAS DE RECOMENDACIÓN**

Durante el presente escrito, se ha enfatizado en la ingente cantidad de datos que se generan diariamente en las organizaciones; cada compra que realiza un cliente lleva implícita una información valiosa para la entidad. El creciente volumen de datos, que es analizado a través de una tecnología en constante desarrollo, permite manejar y procesar decisiones relativas a las organizaciones. Permite conocer la relación que puede ser establecida con el cliente y, en consecuencia, adaptar el resultado de la misma en función de la relación que se ha creado (Liu, Wang & Huang, 2017).

Tal metodología no debe ser entendida como aplicable exclusivamente en tornos empresariales sino también en entidades públicas como una biblioteca universitaria, donde los estudiantes pueden encontrarse limitados en el uso de los recursos que la propia biblioteca dispone por falta de conocimiento sobre los materiales bibliográficos a su disposición (Nzivo & Chuanfu, 2013).

En este sentido, el uso de los datos debe ser contemplado; el exceso de datos puede interferir en los procesos cognitivos de los usuarios y, consiguientemente, obtener una búsqueda de información ineficaz e ineficiente (Yi, Chen & Cong, 2018).

### **10.1. SISTEMA DE RECOMENDACIÓN EN BIBLIOTECAS**

Las bibliotecas digitales son las extensiones de las bibliotecas físicas en el mundo digital que nos envuelve. En este contexto, dichas extensiones digitales amplían el número tanto de recursos como de servicios. Además, las bibliotecas digitalizadas permiten dar una cobertura a una audiencia de usuarios más amplia, considerando este hecho como una ventaja. En definitiva, una biblioteca digital permite a sus usuarios tener de forma accesible contenidos y servicios a través de una aplicación (Porcel, Moreno & Herrera-Viedma, 2009).

De esta manera, para garantizar a los usuarios de las bibliotecas un servicio de valor, se han aplicado sistemas de gestión de bases de datos para analizar aquellos datos que generan los lectores y, como resultado, ofrecerles un servicio mejorado y personalizado. Wu, Lee & Kao (2004) señalan que en una biblioteca no influye el material disponible que posee, sino el material que realmente es usado por los lectores. Por este motivo, la metodología análisis de cesta de la compra debe ser aplicada para conseguir un sistema de recomendación personalizado para consecuentemente aportar a los usuarios información individualizada acerca de las opciones que pueden elegir.

Los sistemas de recomendación son aquellos sistemas que producen recomendaciones individualizadas y guías personalizadas para el usuario. Como refleja Burke (2007) el resultado proporcionado por un sistema de recomendación es entendido como una sugerencia digna de consideración, es el resultado generado por la interpretación de información recuperada acerca de lo que el usuario ha consultado.

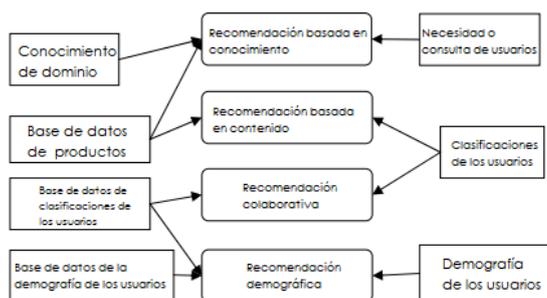
El uso de los sistemas de recomendación ha aumentado en las organizaciones a lo largo de la última década. Una de las explicaciones que argumentan dicho suceso reside en que el número de productos que ofrecen las entidades ha aumentado. Los sistemas de recomendación pues ayudan a predecir el interés del consumidor y le recomiendan productos ajustados a los intereses mostrados. De este modo, existen dos maneras de enfocar los sistemas de recomendación: filtración colaborativa y filtración basada en contenido. Por un lado, la filtración colaborativa hace referencia a la construcción de un modelo a través de los comportamientos pasados de los consumidores, prediciendo pues los productos que pueden suscitar mayor interés en los clientes. Por otro lado, la filtración basada en contenido es el enfoque generado a partir de las características de los elementos seleccionados por los clientes (Rathnavell & Kelkar, 2015).

La circulación de libros que se genera en una biblioteca es considerada una fuente de datos adecuada para diseñar un sistema de recomendación; el sistema de préstamos registra el uso de los libros por parte de los usuarios, concediendo una información valiosa. Destacando que los sistemas de recomendación son un método efectivo para la extracción de información y descubrimiento de contenido para los usuarios, se combinan la búsqueda o consulta de ejemplares con la navegación, dirigida o no, permitiendo a los usuarios del sistema de recomendación una información precisa y satisfactoria en función de sus deseos (Davidson, Liebold, Liu, Nandy, Van Vleet, Gargi, & Sampath, 2010). En definitiva, un sistema de recomendación aplicado al área de bibliotecas permitirá al usuario tener un servicio más satisfactorio, puesto que gracias a la información personalizada que posee el sistema, la búsqueda que realice será considerada rápida y adecuada a los intereses del lector.

## 10.2. TÉCNICAS BASADAS EN CONOCIMIENTOS

Para diseñar un sistema de recomendación se han identificado diversos tipos de técnicas basadas en fuentes de conocimiento, las cuales se reflejan en la Figura 6, y se desarrollan a continuación.

**Figura 6:** Técnicas de recomendación y sus respectivas fuentes de conocimiento



Fuente: Burke (2007)

- Sistema de recomendación con filtrado colaborativo. El sistema genera recomendaciones basándose en preferencias tanto implícitas como explícitas de diversos usuarios, es decir, se estiman las valoraciones de los productos utilizando la opinión de diversos consumidores, asociando los usuarios que presentan gustos similares. En este tipo de sistema de recomendación no se manifiesta una necesidad de información acerca del producto, puesto que se estima en función de las opiniones que han aportado los usuarios del entorno (Good, Schafer, Konstan, Borchers, Sarwar, Herlocker, & Riedl, 1999).
- Sistema de recomendación con filtrado basado en contenido. Este tipo de sistemas genera recomendaciones basándose tanto en los términos que utilizan los usuarios para describir los productos como la información que se posee acerca del cliente. En este contexto, se realiza una recomendación en función de las coincidencias de las descripciones de los productos y los perfiles que presenten los consumidores (Basu, Hirsh & Cohen, 1998).  
Se destaca la eficacia que presenta este tipo de sistema de recomendación, debido a que se basa en características que presenta el producto, en vez de enfocarse en las valoraciones que manifiestan los clientes.
- Sistema de recomendación con filtrado demográfico. Este tercer tipo de sistema de recomendación establece recomendaciones basándose en el perfil demográfico del usuario. La recomendación de productos se genera a través de diversos nichos demográficos (Pazzani, 1999).
- Sistema de recomendación con filtrado basado en conocimiento. Se generan recomendaciones en función de las inferencias de productos, extrayendo pues conocimiento explícito acerca de las características de los productos que satisfacen a los consumidores (Burke, 2002).

Las técnicas que se han mostrado presentan beneficios y desventajas; por este motivo, se puede utilizar un híbrido de técnicas que permitan suavizar aquellas desventajas que presentan los distintos tipos de sistema de recomendación y, consecuentemente, alzar los beneficios que poseen las mismas. Cabe señalar que la tarea de recomendación requiere lo que se denomina *relevante retroalimentación*, la cual es un proceso cíclico que se nutre de la información cedida por el usuario (Porcel, Moreno & Herrera-Viedma, 2009).

## **11. CASO PRÁCTICO 2: SISTEMA DE RECOMENDACIÓN PROPUESTO PARA LA BIBLIOTECA DE LA UNIVERSIDAD DE SEVILLA**

Para verificar el efecto de mejora que produce la aplicación del método análisis cesta de la compra en la configuración un servicio bibliográfico de recomendación personalizada, se ha considerado que la mejor opción para realizar la verificación mencionada es un caso real y cercano al ámbito universitario. Por este motivo, se ha optado por realizar un análisis con los datos proporcionados por la Biblioteca General de la Universidad de Sevilla, creando un sistema de recomendación con filtrado basado en conocimiento, puesto que se van a extraer inferencias de los ejemplares que los usuarios de

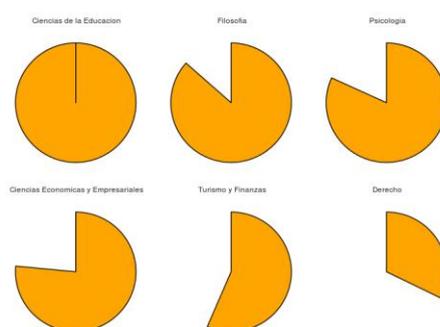
la Biblioteca de la universidad de Sevilla toman como préstamos de la misma; se obtendrá pues un conocimiento explícito de las características de los libros que son prestados.

La finalidad que presenta este segundo caso práctico es estudiar los datos proporcionados por la Biblioteca General de la Universidad de Sevilla, los cuales permiten recomendar a los usuarios material disponible de forma eficaz y, además, que la administración de la Biblioteca conozca el flujo de préstamos y las asociaciones entre ellos. La base de datos que ha sido proporcionada por la Biblioteca General de la Universidad de Sevilla está compuesta por un total de 72.455 registros de préstamos de material bibliográfico. El periodo temporal correspondiente a la extracción de dichos datos se extiende desde el 1 diciembre 2018 hasta el 31 marzo 2019. Para el presente estudio, se ha filtrado el campus universitario Ramón y Cajal, el cual comprende las bibliotecas de las siguientes facultades universitarias: Facultad de Ciencias Económicas y Empresariales, Facultad de Filosofía, Facultad de Psicología, Facultad de Turismo y Finanzas, Facultad de Derecho y Facultad de Ciencias de la Educación.

A continuación, se muestran una serie de figuras exploratorias de los datos cedidos por la Biblioteca General de la Universidad de Sevilla, que permite realizar un estudio sobre los datos que van a ser empleados en la creación del mencionado sistema de recomendación personalizado. Los siguientes datos se encuentran acotados a la población denominada como estudiantes de grado, por lo que se realizará una representación en función de dicha población de forma significativa que guíe las asociaciones que se realizarán para formar el sistema de recomendación.

En la Figura 7 se advierte que el centro con mayor número de préstamos per cápita durante el período estudiado corresponde a la Facultad de Ciencias de la Educación, siendo el Centro con menor número la Facultad de Derecho.

**Figura 7:** % de préstamos per cápita relativizado al caso de la Facultad de Ciencias de la Educación (mayor % de préstamos per cápita).

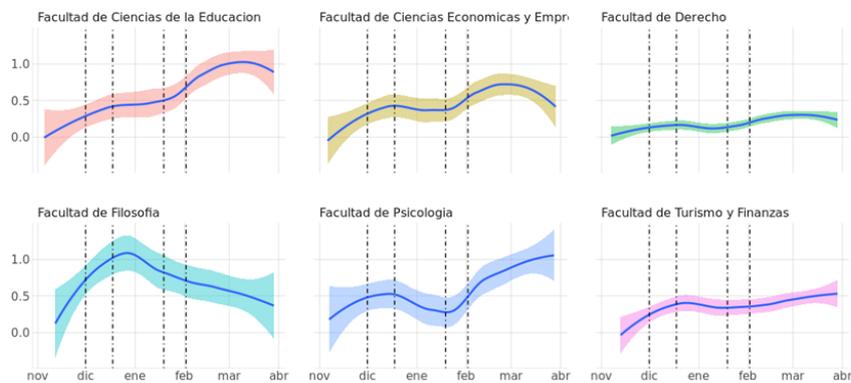


Fuente: Elaboración propia.

En la Figura 8 se observa la evolución de los préstamos per cápita en los distintos centros analizados. Si bien se observa una leve estacionalidad, en general, destaca el caso de la Facultad de Filosofía con un comportamiento más acusado durante el mes previo a la celebración de la primera convocatoria. Asimismo, el comportamiento de la Facultad de Derecho es cuasi-uniforme durante el período analizado con una leve pendiente positiva, situándose el número de préstamos per cápita por debajo de 0,5.

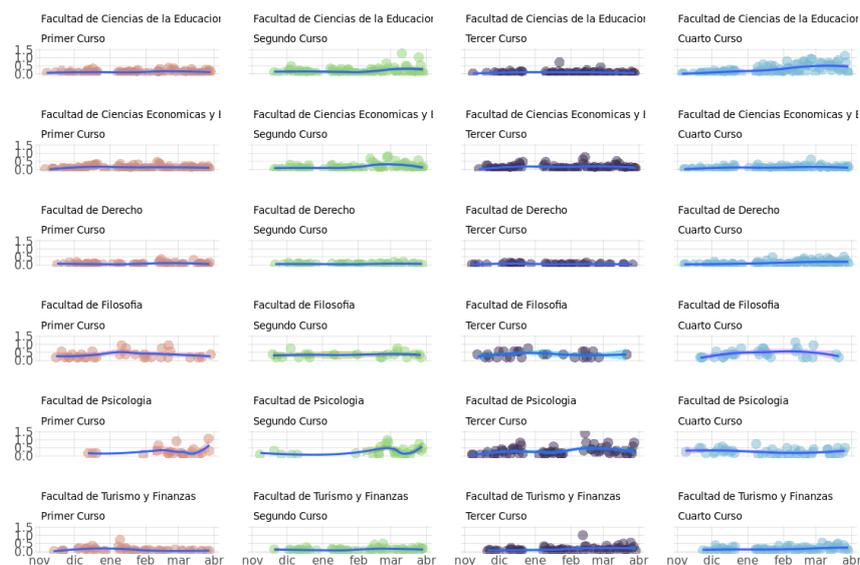
La Figura 9 muestra la evolución por curso y centro de los préstamos concedidos a los estudiantes de grado. Por su parte, en la Figura 10 el porcentaje diario de préstamo de ejemplares en la población de los estudiantes de grado se concentra en su mayoría entre los valores 0 y 1, especialmente es evidenciable el reducido número de préstamos por estudiante de grado (mediana < 0,5) en las Facultades de Turismo y Finanzas, Facultad de Ciencias Económicas y Empresariales y Facultad de Derecho. La Facultad de Ciencias de la Educación y la Facultad de Filosofía, por el contrario, acreditan una mediana de porcentaje de préstamos por alumno superior a 0,5, que en el caso de la Facultad de Ciencias de la Educación supera levemente el valor 3 préstamos / alumno de grado.

**Figura 8:** % diario de préstamos per cápita por centro seleccionado.



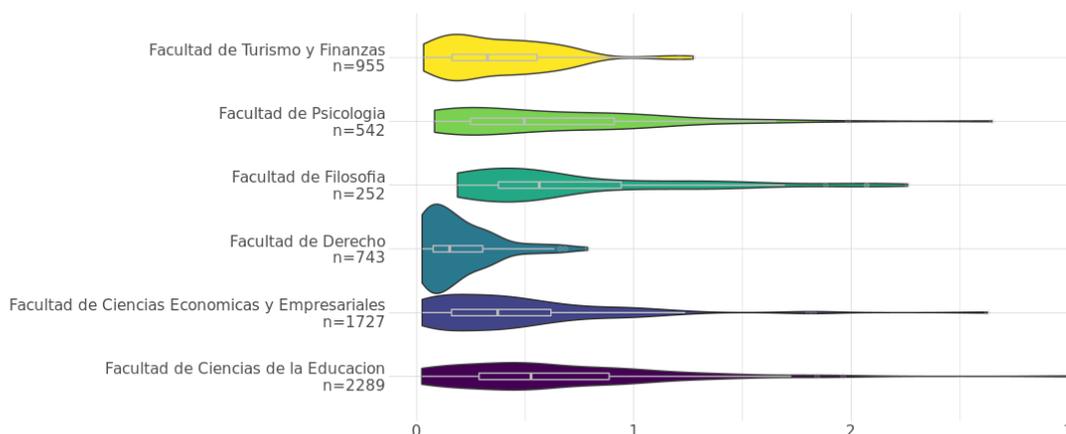
Fuente: Elaboración propia.

**Figura 9:** % diario de préstamos per cápita por curso y centro seleccionado.



Fuente: Elaboración propia.

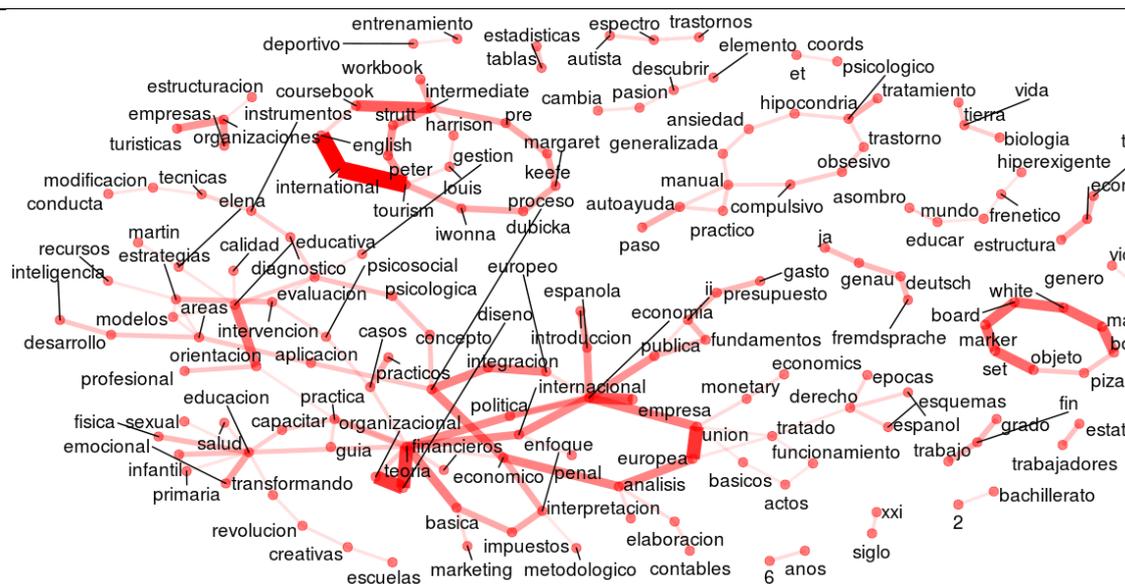
**Figura 10:** Diagrama de violín (y caja) representando el eje relativizado por Centro.



Fuente: Elaboración propia.

En la Figura 11 se han utilizado bigramas para asociar aquellas palabras que más asiduamente se han repetido de forma conjunta. Como ejemplo se representa una fuerte unión entre la serie de palabras *empresa, economía, europea, análisis, teoría, organizacional* entre otras; del mismo modo, otro conjunto de palabras fuertemente asociadas se puede apreciar a la izquierda de la Figura 11, i.e. la relación entre las palabras *coursebook, english, international, intermediate*. El conjunto de palabras que se encuentran unidas permite aventurar la bibliografía que los alumnos de grado buscan en el conjunto de bibliotecas perteneciente al campus universitario Ramón y Cajal.

**Figura 11:** Red de palabras según frecuencia de aparición por bigramas en los títulos de libros prestados.



Fuente: Elaboración propia.

**Figura 12:** % de préstamos de los principales títulos de libros por población y centro.



Fuente: Elaboración propia.

Haciendo referencia a la Figura 12, el libro con porcentaje de préstamos en la Facultad de Ciencias Económicas y Empresariales es *Teoría y diseño organizacional*; por otro lado, *Ética a Nicómaco* es el libro con mayor porcentaje de préstamo por alumno de grado. Con respecto a la Facultad de Psicología el libro *Evaluación psicológica: concepto, proceso y aplicación en las áreas del desarrollo y de la inteligencia* evidencia el porcentaje más alto en el sistema de préstamos de dicho centro. Por último, en la Facultad de Turismo y Finanzas el libro con porcentaje más alto de préstamos pertenece al libro titulado *English for international tourism. Peter Strutt intermediate coursebook*.

A continuación, se ha determinado que los indicadores 'soporte y lift' son los que aportan mayor información; el indicador lift aporta información acerca de la relación que poseen las asociaciones de libros; además, el indicador soporte proporciona la probabilidad de que un conjunto de ejemplares sea elegido simultáneamente.

### 11.1. FACULTAD DE DERECHO

Referenciando la Tabla 31, la asociación destacable según el indicador soporte reside en los ejemplares *Ejecución de sentencias en materia laboral* y *Ejecución en el orden jurisdiccional social: adaptado a los nuevos textos refundidos del estatuto de los trabajadores (rdl 1)*. Como se puede observar el orden de la asociación no es relevante.

Las siguientes asociaciones presentan las mismas propiedades. No obstante, en la asociación número 9 formada por los libros *Organización internacional del trabajo* y *Derechos humanos en la América Latina* y *OIT y Derechos humanos: memoria del director general (parte i) a la conferencia internacional del trabajo, quincuagésima segunda reunión, 1968*, presenta un indicador soporte más bajo, es decir, existe un 0,3 % de probabilidad de adquirirlos conjuntamente.

**Tabla 31:** Tabla de asociaciones de libros pertenecientes a la Biblioteca de la Facultad de Derecho ordenados por el indicador soporte.

	LHS	RHS	SOPORTE	CONFIANZA	LIFT
1	{ejecución de sentencias en materia laboral}	{ejecución en el orden jurisdiccional social: adaptado a los nuevos textos refundidos del estatuto de los trabajadores (rdl 1)}	0,006	1,0	154,0
2	{ejecución en el orden jurisdiccional social: adaptado a los nuevos textos refundidos del estatuto de los trabajadores (rdl 1)}	{ejecución de sentencias en materia laboral}	0,006	1,0	154,0
3	{eutanasia: estudio filosófico-jurídico}	{eutanasia}	0,006	1,0	154,0
4	{eutanasia}	{eutanasia: estudio filosófico-jurídico}	0,006	1,0	154,0
5	{eutanasia: estudio filosófico-jurídico}	{eutanasia hoy: un debate abierto}	0,006	1,0	154,0
6	{eutanasia hoy: un debate abierto}	{eutanasia: estudio filosófico-jurídico}	0,006	1,0	154,0
7	{eutanasia}	{eutanasia hoy: un debate abierto}	0,006	1,0	154,0
8	{eutanasia hoy: un debate abierto}	{eutanasia}	0,006	1,0	154,0
9	{organización internacional del trabajo y los derechos humanos en la América Latina}	{OIT y los derechos humanos: memoria del director general (parte i) a la conferencia internacional del trabajo, quincuagésima segunda reunión, 1968}	0,003	1,0	308,0
10	{OIT y los derechos humanos: memoria del director general (parte i) a la conferencia internacional del trabajo, quincuagésima segunda reunión, 1968}	{organización internacional del trabajo y los derechos humanos en la América latina}	0,003	1,0	308,0
11	{trabajo, mercado de trabajo y relaciones laborales}	{jurisdicción social: comentarios y jurisprudencia: ley 36-2011 de 10 de octubre}	0,003	1,0	308,0

Fuente: Elaboración propia.

Con respecto a la Tabla 32, se puede apreciar que la asociación compuesta por los libros *Organización internacional del trabajo* y los *Derechos Humanos en la América Latina* y *OIT y los Derechos Humanos: memoria del director general (parte i) a la conferencia internacional del trabajo, quincuagésima segunda reunión, 1968* posee un indicador lift elevado. Se presenta por tanto un nivel alto de interdependencia entre estos dos libros, deduciendo que existe una relación positiva entre ambos. Asimismo, el indicador

confianza señala una probabilidad del 100 % de escoger el libro *OIT y los Derechos Humanos: memoria del director general (parte i) a la conferencia internacional del trabajo, quincuagésima segunda reunión, 1968* una vez se ha seleccionado el libro *Organización internacional del trabajo y los Derechos Humanos en la América Latina*, y viceversa.

Reflejado en la Tabla 32, los ejemplares que se muestran presentan un patrón parecido al que se ha analizado, excepto la última asociación, que posee el indicador más bajo recogido en la tabla con un indicador lift mayor que 1. Se determina que la relación existente entre dichos ejemplares es positiva e interdependiente.

**Tabla 32:** Tabla de asociaciones de libros pertenecientes a la Biblioteca de la Facultad de Derecho ordenados por el indicador lift.

	LHS	RHS	SOPORTE	CONFIANZA	LIFT
1	{organización internacional del trabajo y los derechos humanos en la América latina}	{OIT y los derechos humanos: memoria del director general (parte i) a la conferencia internacional del trabajo, quincuagésima segunda reunión, 1968}	0,003	1,0	308,0
2	{ejecución de sentencias en materia laboral}	{ejecución en el orden jurisdiccional social: adaptado a los nuevos textos refundidos del estatuto de los trabajadores (rdl 1)}	0,003	1,0	308,0
3	{trabajo, mercado de trabajo y relaciones laborales}	{jurisdicción social: comentarios y jurisprudencia: ley 36-2011 de 10 de octubre}	0,003	1,0	308,0
4	{jurisdicción social: comentarios y jurisprudencia: ley 36-2011 de 10 de octubre}	{trabajo, mercado de trabajo y relaciones laborales}	0,003	1,0	308,0
5	{estrés y trabajo: como hacerlos compatibles: caracterización, evaluación e intervención de las relaciones disfuncionales entre ambos}	{estrés laboral: guía para empresarios y empleados}	0,003	1,0	308,0
...					
152	{tratamiento jurídico de la inmigración}	{eu charter of fundamental rights: a commentary}	0,003	0,50	154,0
153	{best friends. j. holderness & w. superfine 4 primary. class book}	{estudios sobre el decreto-ley como fuente del derecho financiero y tributario}	0,003	1,0	154,0
154	{estudios sobre el decreto-ley como fuente del	{best friends. j. holderness & w. superfine 4 primary. class book}	0,003	0,50	154,0

	derecho financiero y tributario}				
...					
245	{teoría del derecho: una concepción de la experiencia jurídica}	{economía}	0,003	0,111	1,629

Fuente: Elaboración propia.

La Tabla 33 recoge aquellas asociaciones entre los libros más prestados en función del curso académico que cursa el alumno del Grado. Como se puede apreciar en la Tabla 33, alumnos que cursan el primer curso tienen la probabilidad más alta de adquirir como préstamo el libro *Economía política*, es decir, existe un 0,9 % de probabilidad de que aquellos alumnos que cursan el primer curso consulten el libro mencionado. Sin embargo, en el tercer el libro *Estatuto de los trabajadores* es la asociación con mayor índice de confianza existe una probabilidad del 32,5 % de consultar el libro estatuto de los trabajadores.

**Tabla 33:** Tablas de asociaciones según el curso del alumno y el libro más prestado en la Facultad de Derecho.

	LHS	RHS	SOPORTE	CONFIANZA	LIFT
1	{Primer Curso}	{economía política}	0,009	0,247	11,563
2	{Tercer Curso}	{estatuto de los trabajadores}	0,006	0,325	25,895
3	{Primer Curso}	{economía}	0,006	0,180	9,736
4	{Segundo Curso}	{estatuto de los trabajadores}	0,003	0,112	8,935
5	{Segundo Curso}	{economía laboral}	0,002	0,102	25,324
6	{Quinto Curso}	{transporte multimodal: concepto y sujetos}	0,002	0,170	89,766
7	{Quinto Curso}	{tratado de derecho penal internacional}	0,001	0,106	22,441
8	{Quinto Curso}	{tablas estadísticas}	0,001	0,106	7,358

Fuente: Elaboración propia.

## 11.2. FACULTAD DE CIENCIAS ECONÓMICAS Y EMPRESARIALES

Haciendo referencia a la Tabla 34, se observa que la mayoría de las asociaciones presentan un indicador lift elevado. Resulta interesante resaltar la asociación 231 y 232; ambas poseen el mismo indicador lift y el mismo indicador soporte, sin embargo, a pesar de que ambas asociaciones están formadas por los ejemplares *Transparencia, acceso a la información pública* y *buen gobierno: ley 19* y *Estudio sistemático de la ley de contratos del sector público*, el indicador confianza difiere; este hecho se traduce en que si se selecciona primero el préstamo del libro *Estudio sistemático de la ley de contratos del sector público* existe un 50 % menos de probabilidad de adquirir el préstamo del ejemplar *Transparencia, acceso a la información pública* y *buen gobierno: ley 19*.

En la Tabla 34 también se muestra la asociación 673; está compuesta por los libros *Teoría básica de marketing* y *Economía de la empresa*, siendo la asociación menos representativa de la tabla.

**Tabla 34:** Tabla de asociaciones de libros pertenecientes a la Biblioteca de la Facultad de Ciencias Económicas y Empresariales ordenados por el indicador lift.

	LHS	RHS	SOPORTE	CONFIANZA	LIFT
1	{trampa de la deuda: tercer mundo y dependencia}	{economía para un mundo en desarrollo: introducción a los principios, problemas y políticas para el desarrollo}	0,002	1,0	614,0
2	{economía para un mundo en desarrollo: introducción a los principios, problemas y políticas para el desarrollo}	{trampa de la deuda: tercer mundo y dependencia}	0,002	1,0	614,0
3	{test sociométrico}	{terapias de tercera generación como terapias contextuales}	0,002	1,0	614,0
4	{terapias de tercera generación como terapias contextuales}	{test sociométrico}	0,002	1,0	614,0
5	{eco innovación como clave para el éxito empresarial: tendencias, beneficios y primeros pasos para eco innovar}	{rebranding verde}	0,002	1,0	614,0
6	{rebranding verde}	{eco innovación como clave para el éxito empresarial: tendencias, beneficios y primeros pasos para eco innovar}	0,002	1,0	614,0
...			0,002		
231	{transparencia, acceso a la información pública y buen gobierno: ley 19}	{estudio sistemático de la ley de contratos del sector público}	0,002	1,0	307,0
232	{estudio sistemático de la ley de contratos del sector público}	{transparencia, acceso a la información pública y buen gobierno: ley 19}	0,002	0,50	307,0
233	{teoría de la decisión y de los juegos}	{optimización: programación matemática y aplicaciones a la economía}	0,002	1,0	307,0
234	{optimización: programación matemática y	{teoría de la decisión y de los juegos}	0,002	0,50	307,0

	aplicaciones a la economía}				
...					
673	{teoría básica de marketing}	{economía de la empresa}	0,005	0,111	1,083

Fuente: Elaboración propia.

En la Tabla 35 se muestran las asociaciones ordenadas de mayor a menor en función del indicador soporte. La asociación formada por los libros *Economía pública II* y *Teoría básica de los impuestos: un enfoque económico* obtiene los valores más altos para los tres indicadores expuestos. Cuando se selecciona el libro *Economía pública II* existe una probabilidad de 2,4 % de seleccionar conjuntamente el ejemplar *Teoría básica de los impuestos: un enfoque económico*. Existe una probabilidad de 78,9 % de escoger *Teoría básica de los impuestos: un enfoque económico* habiendo sido prestado el libro *Economía pública II*. Por el contrario, cuando el ejemplar *Teoría básica de los impuestos: un enfoque económico* aparece en el lado izquierdo (LHS) de la asociación, la probabilidad de escoger como préstamo *economía pública II* queda reducida a un 48,3 % de probabilidad.

La Tabla 36 alude a los préstamos más usuales en función del curso que está realizando el usuario. La probabilidad de pertenecer al segundo curso y pedir el préstamo del ejemplar *Teoría y diseño organizacional* evidencia la mayor probabilidad de la tabla, 4 % de probabilidad. La asociación del primer curso relacionada con el libro *Economía de la empresa* adquiere mayor índice de confianza y lift, determinando pues que si se está cursando primero de grado existe una probabilidad de 35,3 % de tomar prestado de la Biblioteca el ejemplar *Economía de la empresa*.

**Tabla 35:** Tabla de asociaciones de libros pertenecientes a la Biblioteca de la Facultad de Ciencias Económicas y Empresariales ordenados por el indicador soporte.

	LHS	RHS	SOPORTE	CONFIANZA	LIFT
1	{economía pública II}	{teoría básica de los impuestos: un enfoque económico}	0,024	0,789	15,636
2	{teoría básica de los impuestos: un enfoque económico}	{economía pública II}	0,024	0,483	15,636
3	{teoría básica de los impuestos: un enfoque económico}	{economía internacional: teoría y política}	0,024	0,483	5,825
4	{economía internacional: teoría y política}	{teoría básica de los impuestos: un enfoque económico}	0,024	0,294	5,825
5	{economía pública i: i. fundamentos, II. presupuesto y gasto}	{economía española: una introducción}	0,023	0,424	6,512
6	{economía española: una introducción}	{economía pública i: i. fundamentos, II. presupuesto y gasto}	0,023	0,350	6,512
7	{economía pública i: i. fundamentos, II. presupuesto y gasto}	{teoría básica de los impuestos: un enfoque económico}	0,021	0,394	7,802

8	{teoría básica de los impuestos: un enfoque económico}	{economía pública i: i. fundamentos, II. presupuesto y gasto}	0,021	0,419	7,802
9	{economía española: una introducción}	{teoría básica de los impuestos: un enfoque económico}	0,021	0,325	6,437
10	{teoría básica de los impuestos: un enfoque económico}	{economía española: una introducción}	0,021	0,419	6,437

Fuente: Elaboración propia.

**Tabla 36:** Tablas de asociaciones según el curso del alumno y el libro más prestado en la Facultad de Ciencias Económicas y Empresariales.

	LHS	RHS	SOPORTE	CONFIANZA	LIFT
1	{Segundo Curso}	{teoría y diseño organizacional}	0,040	0,324	7,053
2	{ Primer Curso}	{economía de la empresa}	0,031	0,353	9,554
3	{ Tercer Curso}	{economía española: una introducción}	0,019	0,188	8,632
4	{ Segundo Curso}	{ economía de la unión europea: análisis económico del proceso de integración europeo}	0,019	0,154	5,475
5	{ Tercer Curso}	{ teoría básica de los impuestos: un enfoque económico}	0,018	0,179	8,982
6	{ Tercer Curso}	{ economía internacional: teoría y política}	0,016	0,155	5,956
7	{ Tercer Curso}	{ economía pública i: i, fundamentos, II, presupuesto y gasto}	0,011	0,113	6,711
8	{ Primer Curso}	{ economía política}	0,010	0,120	5,605
9	{ Primer Curso}	{ teoría básica de marketing}	0,009	0,101	11,163
10	{ Cuarto Curso}	{ economics of monetary union}	0,008	0,110	12,899

Fuente: Elaboración propia.

### 11.3. FACULTAD DE CIENCIAS DE LA EDUCACIÓN

En la Tabla 37 se muestra que la asociación formada por los ejemplares *Tecnología educativa: la formación del profesorado en la era de internet* y *Evaluación y gestión de la calidad educativa: un enfoque metodológico* obtiene el indicador soporte más alto; evidencian una probabilidad de 1 % de adquirirse simultáneamente.

Los ejemplares *Jugar a pensar: recursos para aprender a pensar en educación infantil* y *Juegos para la educación infantil. Sofia Kloppe, Jorge Batllor ; [ilustraciones] Horacio Elena preescolar* presentan una de las asociaciones más positivas de la tabla, además de poseer una confianza del 100 %.

**Tabla 37:** Tabla de asociaciones de libros pertenecientes a la Biblioteca de la Facultad de Ciencias de la Educación ordenados por el indicador soporte.

	LHS	RHS	SOPORTE	CONFIANZA	LIFT
1	{tecnología educativa: la formación del profesorado en la era de internet}	{evaluación y gestión de la calidad educativa: un enfoque metodológico}	0,010	0,223	2,035
2	{evaluación y gestión de la calidad educativa: un enfoque metodológico}	{tecnología educativa: la formación del profesorado en la era de internet}	0,007	0,375	25,469
3	{teoría y práctica de la orientación educativa}	{orientación educativa. Elena Martín, Isabel Sole (coords.)... [et al.] modelos y estrategias de intervención}	0,007	0,667	30,185
4	{orientación educativa. Elena Martín, Isabel Sole (coords.)... [et al.] modelos y estrategias de intervención}	{teoría y práctica de la orientación educativa}	0,007	0,334	30,185
...					
246	{educar en verde: ideas para acercar a niños y niñas a la naturaleza}	{educación al aire libre}	0,001	0,3334	271,667
247	{beso, beso!}	{Julieta y su caja de colores}	0,001	1,0	271,667
248	{Julieta y su caja de colores}	{beso, beso!}	0,001	0,334	271,667
249	{jugar a pensar: recursos para aprender a pensar en educación infantil}	{juegos para la educación infantil. Sofia Kloppe, Jorge Batllori ; [ilustraciones] Horacio Elena preescolar}	0,001	1,0	815,0
250	{juegos para la educación infantil. Sofia Kloppe, Jorge Batllori ; [ilustraciones] Horacio Elena preescolar}	{jugar a pensar: recursos para aprender a pensar en educación infantil}	0,001	1,0	815,0

Fuente: Elaboración propia.

#### 11.4. FACULTAD DE TURISMO Y FINANZAS

Observando la Tabla 38 se evidencia un nivel elevado de interdependencia entre los ejemplares. La asociación de los ejemplares *Estadística aplicada: economía y ciencias sociales* y *Elementos básicos de estadística económica y empresarial*, muestra un valor confianza 50 % cuando se adquiere primero el libro *Estadística aplicada: economía y ciencias sociales*. De la misma manera ocurre con los ejemplares *Estructuración de las organizaciones* y *Teoría de la organización: un enfoque estratégico*.

**Tabla 38:** Tabla de asociaciones de libros pertenecientes a la Biblioteca de la Facultad de Turismo y Finanzas ordenados por el indicador lift.

	LHS	RHS	SOPORTE	CONFIANZA	LIFT
1	{turismo y género: la perspectiva inclusiva en la gestión turística}	{turismo y genero}	0,002	1,0	368,0
2	{turismo y genero}	{turismo y género: la perspectiva inclusiva en la gestión turística}	0,002	1,0	368,0
3	{organizaciones no gubernamentales: un modelo integral de gestión y control}	{ONG: un sistema de indicadores para su evaluación y gestión}	0,002	1,0	368,0
4	{ONG: un sistema de indicadores para su evaluación y gestión}	{organizaciones no gubernamentales: un modelo integral de gestión y control}	0,002	1,0	368,0
5	{eventos de moda y su influencia en el sector turístico: trabajo fin de grado: grado en turismo}	{turismo de eventos propuesta de estudio del perfil del turista de festivales de música, trabajo fin de grado: grado en turismo}	0,002	1,0	368,0
...					
124	{elementos básicos de estadística económica y empresarial}	{estadística aplicada: economía y ciencias sociales}	0,002	1,0	184,0
125	{estadística aplicada: economía y ciencias sociales}	{elementos básicos de estadística económica y empresarial}	0,002	0,50	184,0
126	{teoría de la organización: un enfoque estratégico}	{estructuración de las organizaciones}	0,002	1,0	184,0
127	{estructuración de las organizaciones}	{teoría de la organización: un enfoque estratégico}	0,002	0,50	184,0

Fuente: Elaboración propia.

Cuando se adquiere el libro *English for international tourism. Louis Harrison intermediate workbook* existe una probabilidad del 5,1 % (véase Tabla 39) de tomar prestado conjuntamente con el libro *English for international tourism. Peter Strutt intermediate coursebook*.

En las cinco primeras asociaciones de la Tabla 39 se aprecia que la asociación entre *English for international tourism: intermediate workbook with key* y *English for international tourism. Peter Strutt intermediate coursebook* posee el valor más elevado en el indicador confianza, cuando se ha tomado prestado el libro *English for international tourism: intermediate workbook with key*, existe una probabilidad del 100 % de seleccionar también como préstamo el libro *English for international tourism. Peter Strutt intermediate coursebook*.

**Tabla 39:** Tabla de asociaciones de libros pertenecientes a la Biblioteca de la Facultad de Turismo y Finanzas ordenados por el indicador soporte.

	LHS	RHS	SOPORTE	CONFIANZA	LIFT
1	{english for international tourism. Louis Harrison intermediate workbook}	{english for international tourism. Peter Strutt intermediate coursebook}	0,051	0,904	5,945
2	{english for international tourism. Peter Strutt intermediate coursebook}	{english for international tourism. Louis Harrison intermediate workbook}	0,051	0,339	5,945
3	{english for international tourism: intermediate workbook with key}	{english for international tourism. Peter Strutt intermediate coursebook}	0,032	1,0	6,571
4	{english for international tourism. Peter Strutt intermediate coursebook}	{english for international tourism: intermediate workbook with key}	0,032	0,214	6,571
5	{english for international tourism. iwonna dubicka y margaret o'keefe pre-intermediate workbook}	{english for international tourism. iwonna dubicka y margaret o'keefe pre-intermediate coursebook}	0,029	0,785	13,768
...					
246	{enciclopedia del community manager}	{english for international tourism. Louis Harrison intermediate workbook}	0,002	1,0	17,523
247	{enciclopedia del community manager}	{english for international tourism. Peter Strutt intermediate coursebook}	0,002	1,0	17,523
248	{estados contables: elaboración, análisis e interpretación}	{estados financieros: teoría y casos prácticos}	0,002	1,0	6,571
249	{entusiasmo: precariedad y trabajo creativo en la era digital}	{turismo idiomático :grado de importancia del castellano en la elección de España como destino para los estudiantes extranjeros: trabajo fin de grado: grado en turismo}	0,002	0,50	92,0
250	{turismo idiomático :grado de importancia del castellano en la elección de España como destino para los estudiantes extranjeros: trabajo fin de grado: grado en turismo}	{entusiasmo: precariedad y trabajo creativo en la era digital}	0,002	0,50	92,0

Fuente: Elaboración propia.

Representando en la Tabla 40 las asociaciones del curso a que pertenece el alumno, se observa que cuando se cursa primer curso existe una probabilidad de 1,8 % de solicitar un préstamo del libro *Empresas y organizaciones turísticas*. Del mismo modo, para el segundo curso el libro con mayor probabilidad de ser adquirido es *Estructura económica del turismo* y para el tercer curso el libro titulado *English for international tourism, Peter Strutt intermediate coursebook*. Por último, el cuarto curso presenta el menor valor para el indicador soporte, 0,8 %; cuando se cursa cuarto curso existe una probabilidad de seleccionar como préstamo el libro *English for international tourism, Peter Strutt intermediate coursebook*.

**Tabla 40:** Tablas de asociaciones según el curso del alumno y el libro más prestado en la Facultad de Turismo y Finanzas.

	LHS	RHS	SO- PORTE	CON- FIANZA	LIFT
1	{Primer Curso}	{empresas y organizaciones turísticas}	0,018	0,497	20,751
2	{Segundo Curso}	{estructura económica del turismo}	0,014	0,296	17,563
3	{Tercer Curso}	{english for international tourism, Peter Strutt intermediate coursebook}	0,013	0,202	8,047
4	{Tercer Curso}	{english for international tourism, iwonna dubicka y margaret o'keefe pre-intermediate coursebook}	0,010	0,148	12,245
5	{Tercer Curso}	{ja, genau!, deutsch als fremdsprache}	0,009	0,130	8,063
6	{Cuarto Curso}	{english for international tourism, Peter Strutt intermediate coursebook}	0,008	0,109	4,353
7	{Tercer Curso}	{tablas estadísticas}	0,008	0,116	7,990
8	{Segundo Curso}	{ ja, genau!, deutsch als fremdsprache}	0,006	0,128	7,947
9	{Primer Curso}	{turismo en el mundo actual: desafíos de la globalización}	0,004	0,110	27,219
10	{Primer Curso}	{empresas y organizaciones turísticas}	0,018	0,497	20,751

Fuente: Elaboración propia.

## 11.5. FACULTAD DE PSICOLOGÍA

Los libros *Envejecimiento y psicología de la salud* y *Enfermedad de alzhéimer: clínica, tratamiento y rehabilitación* poseen una probabilidad de ser adquiridos conjuntamente del 0,463 %, y un índice de confianza del 100 % de ser seleccionado *Enfermedad de alzhéimer: clínica, tratamiento y rehabilitación* habiendo sido prestado *Envejecimiento y psicología de la salud*, y viceversa.

Más aún, cuando se toma prestado el libro *Test del dibujo de la persona bajo la lluvia: una guía interpretativa* existe un 50 % de probabilidad de adquirir también *Test de goodenough: revisión, ampliación y actualización*. Si se selecciona primero libro *Test de goodenough: revisión, ampliación y actualización* el préstamo más probable es el libro *Test del dibujo de la persona bajo la lluvia: una guía interpretativa*.

**Tabla 41:** Tabla de asociaciones de libros pertenecientes a la Biblioteca de la Facultad de Psicología ordenados por el indicador lift.

	LHS	RHS	SOPORTE	CONFIANZA	LIFT
1	{envejecimiento y psicología de la salud}	{enfermedad de alzheimer: clínica, tratamiento y rehabilitación}	0,004	1,0	216,0
2	{enfermedad de alzheimer: clínica, tratamiento y rehabilitación}	{envejecimiento y psicología de la salud}	0,004	1,0	216,0
3	{educación sexual}	{transitar la cultura: niños y niñas inmigrantes en la educación obligatoria: estudios de caso, multicaso}	0,004	1,0	216,0
4	{transitar la cultura: niños y niñas inmigrantes en la educación obligatoria: estudios de caso, multicaso}	{educación sexual}	0,004	1,0	216,0
...					
96	{test del dibujo de la persona bajo la lluvia: una guía interpretativa}	{técnicas de relajación: manual práctico para adultos, niños y educación especial}	0,004	0,50	108,0
97	{test de goodenough: revisión, ampliación y actualización}	{test del dibujo de la persona bajo la lluvia: una guía interpretativa}	0,004	1,0	108,0
98	{test del dibujo de la persona bajo la lluvia: una guía interpretativa}	{test de goodenough: revisión, ampliación y actualización}	0,004	0,50	108,0
99	{escuela, diversidad cultural e inclusión}	{tratamiento del TOC en niños y adolescentes}	0,004	1,0	108,0
100	{tratamiento del TOC en niños y adolescentes}	{escuela, diversidad cultural e inclusión}	0,004	0,50	108,0

Fuente: Elaboración propia.

La asociación formada por los ejemplares *Tratamiento psicológico de la hipocondría y la ansiedad generalizada: un manual práctico de autoayuda paso a paso* y *Tratamiento psicológico del trastorno obsesivo-compulsivo: un manual de autoayuda paso a paso*, posee el valor más elevado para el indicador soporte, es decir, un 6,018 %. La probabilidad de adquirir el libro *Tratamiento psicológico de la hipocondría y la ansiedad generalizada: un manual práctico de autoayuda paso a paso* una vez ha sido seleccionado como préstamo el ejemplar *Tratamiento psicológico del trastorno obsesivo-compulsivo: un manual de autoayuda paso a paso* es menor en comparación con seleccionar primero el libro *Tratamiento psicológico de la hipocondría y la ansiedad generalizada: un manual práctico de autoayuda paso a paso*.

**Tabla 42:** Tabla de asociaciones de libros pertenecientes a la Biblioteca de la Facultad de Psicología ordenados por el indicador soporte.

	LHS	RHS	SOPORTE	CONFIANZA	LIFT
1	{tratamiento psicológico de la hipocondría y la ansiedad generalizada: un manual práctico de autoayuda paso a paso}	{tratamiento psicológico del trastorno obsesivo-compulsivo: un manual de autoayuda paso a paso}	0,060	0,419	2,448
2	{tratamiento psicológico del trastorno obsesivo-compulsivo: un manual de autoayuda paso a paso}	{tratamiento psicológico de la hipocondría y la ansiedad generalizada: un manual práctico de autoayuda paso a paso}	0,060	0,351	2,448
3	{evaluación psicológica: concepto, proceso y aplicación en las áreas del desarrollo y de la inteligencia}	{estrategias de intervención psicosocial: casos prácticos}	0,046	0,357	2,967
4	{estrategias de intervención psicosocial: casos prácticos}	{evaluación psicológica: concepto, proceso y aplicación en las áreas del desarrollo y de la inteligencia}	0,046	0,384	2,967
5	{estrategias de intervención psicosocial: casos prácticos}	{tratamiento psicológico del trastorno obsesivo-compulsivo: un manual de autoayuda paso a paso}	0,037	0,307	1,796
...					
236	{técnicas de modificación de conducta, tratamiento psicológico de la hipocondría y la ansiedad generalizada: un manual práctico de autoayuda paso a paso}	{tratamiento psicológico del trastorno obsesivo-compulsivo: un manual de autoayuda paso a paso}	0,004	0,50	2,918
237	{estrategias de intervención psicosocial: casos prácticos, terapia de conducta en la infancia: guía de intervención}	{tratamiento psicológico de la hipocondría y la ansiedad generalizada: un manual práctico de autoayuda paso a paso}	0,004	1,0	6,967
238	{terapia de conducta en la infancia: guía de intervención, tratamiento psicológico de la hipocondría y la ansiedad generalizada: un	{estrategias de intervención psicosocial: casos prácticos}	0,004	1,0	8,307

	manual práctico de autoayuda paso a paso}				
239	{estrategias de intervención psicosocial: casos prácticos, tratamiento psicológico de la hipocondría y la ansiedad generalizada: un manual práctico de autoayuda paso a paso}	{terapia de conducta en la infancia: guía de intervención}	0,004	0,142	2,805
240	{estrategias de intervención psicosocial: casos prácticos, evaluación psicológica: concepto, proceso y aplicación en las áreas del desarrollo y de la inteligencia, tratamiento psicológico de la hipocondría y la ansiedad generalizada: un manual práctico de autoayuda paso a paso}	{técnicas de modificación de conducta}	0,004	0,50	13,50

Fuente: Elaboración propia.

En la Tabla 43 se observa que el libro más prestado en el primer curso es *Trabajando con grupos: técnicas de intervención*, mientras que en segundo curso el libro titulado *Escuelas creativas: la revolución que está transformando la educación* adquiere el mayor número de préstamos. Por último, en el tercer curso el libro *Evaluación psicológica: concepto, proceso y aplicación en las áreas del desarrollo y de la inteligencia* es el que posee la mayor probabilidad de ser solicitado.

**Tabla 43:** Tablas de asociaciones según el curso del alumno y el libro más prestado en la Facultad de Psicología.

	LHS	RHS	SO- PORTE	CON- FIANZA	LIFT
1	{Tercer Curso}	{evaluación psicológica: concepto, proceso y aplicación en las áreas del desarrollo y de la inteligencia}	0,012	0,228	16,88 1
2	{Tercer Curso}	{tratamiento psicológico del trastorno obsesivo-compulsivo: un manual de autoayuda paso a paso}	0,008	0,154	16,60 7
3	{Tercer Curso}	{tratamiento psicológico de la hipocondría y la ansiedad generalizada: un manual práctico de autoayuda paso a paso}	0,008	0,149	16,55 7

4	{Tercer Curso}	{estrategias de intervención psico-social: casos prácticos}	0,008	0,145	17,44 7
5	{Segundo Curso}	{escuelas creativas: la revolución que está transformando la educación}	0,006	0,248	41,77 2
6	{Primer Curso}	{trabajando con grupos: técnicas de intervención}	0,005	0,319	58,59 7
7	{Segundo Curso}	{elemento: descubrir tu pasión lo cambia todo}	0,005	0,198	28,80 8
8	{Primer Curso}	{emoción, afecto y motivación: un enfoque de procesos}	0,003	0,153	58,59 7

Fuente: Elaboración propia.

### 11.6. FACULTAD DE FILOSOFÍA

Las asociaciones de la Tabla 44 no muestran diferencias relevantes entre sí, el hecho más destacable es el cambio en el indicador lift, cuando a partir de la asociación número 92 desciende significativamente hasta el valor 39,50.

La Tabla 45 refleja que la asociación con mayor indicador soporte reside en los ejemplares *Ética a Nicómaco* y *Existencialismo es un humanismo*, siendo el valor del índice soporte 3,7 %. Además, la probabilidad de seleccionar como préstamo *Ética a Nicómaco* una vez ha seleccionado *Existencialismo es un humanismo*, y viceversa, es igual a 0,50.

**Tabla 44:** Tabla de asociaciones de libros pertenecientes a la Biblioteca de la Facultad de Filosofía ordenados por el indicador lift

	LHS	RHS	SOPORTE	CON-FIANZA	LIFT
1	{Europa heautentimorume, es decir, que miseramente a si misma se atormenta y lamenta su propia desgracia}	{Judith Butler en disputa: lecturas sobre la performatividad}	0,012	1,0	79,0
2	{Judith Butler en disputa: lecturas sobre la performatividad}	{Europa heautentimorume, es decir, que miseramente a si misma se atormenta y lamenta su propia desgracia}	0,012	1,0	79,0
3	{breve historia y antología de la estética}	{ejercicios hipopresivos: mucho más que abdominales: reprograma tu cuerpo de forma inteligente, alivia tus dolores de espalda, reduce la cintura, mejora tu rendimiento deportivo y sexual}	0,012	1,0	79,0
4	{ejercicios hipopresivos: mucho más que abdominales: reprograma tu cuerpo de	{breve historia y antología de la estética}	0,012	1,0	79,0

	forma inteligente, alivia tus dolores de espalda, reduce la cintura, mejora tu rendimiento deportivo y sexual}				
...					
92	{elogio de lo cotidiano}	{estética de lo feo}	0,012	1,0	39,50
93	{estética de lo feo}	{elogio de lo cotidiano}	0,012	0,50	39,50
94	{trágico}	{bosques de la noche: (poemas, canciones y epigramas)}	0,012	1,0	39,50
95	{bosques de la noche: (poemas, canciones y epigramas)}	{trágico}	0,012	0,50	39,50
96	{belleza múltiple}	{estética de lo feo}	0,012	1,0	39,50

Fuente: Elaboración propia.

**Tabla 45:** Tabla de asociaciones de libros pertenecientes a la Biblioteca de la Facultad de Filosofía ordenados por el indicador soporte.

	LHS	RHS	SOPORTE	CONFIANZA	LIFT
1	{ética a Nicómaco}	{existencialismo es un humanismo}	0,037	0,5	6,583
2	{existencialismo es un humanismo}	{ética a Nicómaco}	0,037	0,5	6,583
3	{obras completas del pseudo Dionisio areopagita}	{ética demostrada según el orden geométrico}	0,025	1,0	19,750
4	{ética demostrada según el orden geométrico}	{obras completas del pseudo Dionisio areopagita}	0,025	0,5	19,750
5	{ecce homo: como se llega a ser lo que se es}	{obras completas}	0,025	1,0	19,750
...					
204	{existencialismos: claves para su comprensión, obras completas}	{existencialismo es un humanismo}	0,012	1,0	13,166
205	{existencialismo es un humanismo, existencialismos: claves para su comprensión}	{obras completas}	0,012	0,5	9,875
206	{existencialismo es un humanismo, obras completas}	{existencialismos: claves para su comprensión}	0,012	0,5	13,166
207	{ética a Nicómaco, existencialismos: claves para su comprensión}	{existencialismo es un humanismo}	0,012	1,0	13,166
208	{ética a Nicómaco, obras completas}	{existencialismo es un humanismo}	0,012	1,0	13,166

Fuente: Elaboración propia.

En la Tabla 46 se evidencia que el libro elegido como préstamo cuando se está cursando el primer curso es *Ética a Nicómaco*. Cuando se cursa el segundo curso se comprueba una probabilidad de 0,3 % de adquirir como préstamo el ejemplar *Ética demostrada según el orden geométrico*. Por último, se destaca que la probabilidad de consultar el libro *Estrategia de Ulises o ética para una sociedad tecnológica* en cuarto curso es de 0,4 %.

**Tabla 46:** Tablas de asociaciones según el curso del alumno y el libro más prestado en la Facultad de Filosofía.

	LHS	RHS	SOPORTE	CONFIANZA	LIFT
1	{Primer Curso}	{ética a Nicómaco}	0,004	0,239	53,027
2	{Cuarto Curso}	{estrategia de Ulises o ética para una sociedad tecnológica}	0,004	0,160	44,883
3	{Cuarto Curso}	{teoría feminista: de la ilustración a la globalización}	0,003	0,128	44,883
4	{Segundo Curso}	{ética demostrada según el orden geométrico}	0,003	0,275	105,475
5	{Primer Curso}	{existencialismo es un humanismo}	0,002	0,149	62,970
6	{Primer Curso}	{Tao Te King: libro del curso y de la virtud}	0,002	0,104	62,970
7	{Segundo Curso}	{breve historia de la filosofía medieval}	0,001	0,10	105,475
8	{Segundo Curso}	{obras completas del pseudo Dionisio areopagita}	0,001	0,10	105,475
9	{Primer Curso}	{ética a Nicómaco}	0,004	0,239	53,027
10	{Cuarto Curso}	{estrategia de Ulises o ética para una sociedad tecnológica}	0,004	0,160	44,883

Fuente: Elaboración propia.

## 12. CONCLUSIONES

En el presente escrito se ha realizado una revisión teórica que ilustra el proceso de recogida de datos y la depuración de los mismos para extraer información valiosa. De este modo, se han expuesto diversas propuestas que permiten a los investigadores conseguir dicho fin, centrándose en el método perteneciente a la minería de datos, reglas de asociación, y desembocando en la técnica análisis cesta de la compra, la cual ha sido aplicada a dos casos prácticos.

Se ha podido reflejar que la utilidad presentada por el análisis cesta de la compra es elevada para el área del marketing. Una de las razones que sustentan su utilidad reside en que permite a los investigadores evaluar la presencia de asociaciones mediante un enfoque inductivo. El análisis cesta de la compra es una técnica adecuada para indicar a los investigadores relaciones de contingencia sin recaer en *consideraciones insostenibles*, además de dar la posibilidad de diseñar propuestas dinámicas a través de datos que se habían inicialmente despreciado por inservibles o *ruidosos*. En general, el análisis

cesta de la compra permite a las organizaciones ofrecer recomendaciones significativas en actuaciones empresariales.

En el estudio de las transacciones recogidas en un supermercado se ha demostrado, a través de la muestra empleada, que la técnica análisis cesta de la compra es una herramienta eficaz para identificar conjuntos de categorías de productos frecuentes en las transacciones registradas; siendo la información obtenida válida para estudiar el comportamiento del consumidor en el punto de venta que maximice a su vez su valor percibido. Se refuerza con esta idea la necesidad de una coordinación en las estrategias de marketing implantadas en las organizaciones a través de la interrelación de las categorías de productos.

Asimismo, se diseña en fase inicial un sistema de recomendación para la Biblioteca General de la Universidad de Sevilla. La influencia digital se ha implantado en las bibliotecas, proporcionando un nuevo modo de acceso. La digitalización permite obtener datos del usuario y, en consecuencia, definir un perfil. Partiendo de él se ha desarrollado un sistema de recomendación a través de la información explícita recogida en los préstamos de libros que realizan los usuarios, ofreciendo de esta manera un servicio personalizado.

Aludiendo a la parte práctica del sistema de recomendación se han realizado asociaciones en función de la facultad para dar un servicio de valor a los estudiantes de cada centro, proporcionando una serie de ejemplares que presentan interdependencia. Adicionalmente, se ha obtenido el libro con más préstamos en función del curso y la facultad del alumno. Estas recomendaciones son válidas tanto para el alumnado, al cual le resultará más eficaz la búsqueda de libros relacionados, y también para la administración de la Biblioteca General de la Universidad de Sevilla, para conocer cuáles son los ejemplares más demandados.

En conclusión, en el presente escrito se ha desarrollado una propuesta basada en el análisis cesta de la compra, que permite a partir de una serie de transacciones recogidas en la base de datos de la organización, proporcionar patrones de compra para utilizarlos en la optimización de las acciones de marketing. Más aún, se ha comprobado la validez de la propuesta en diversas áreas; siendo la finalidad buscada la misma para ambas, encontrar patrones que permitan mejorar la satisfacción percibida por el consumidor y su compromiso con la organización.

En términos de futuras investigaciones se propone la integración en el análisis de elementos descriptivos del usuario y de sus percepciones del servicio.

### 13. REFERENCIAS BIBLIOGRÁFICAS

- Aguinis, H., Beaty, J. C., Boik, R. J., & Pierce, C. A. (2005). Effect size and power in assessing moderating effects of categorical variables using multiple regression: a 30-year review. *Journal of Applied Psychology*, 90(1), 94.
- Aguinis, H., Forcum, L. E., & Joo, H. (2013). Using market basket analysis in management research. *Journal of Management*, 39(7), 1799-1824.
- Agrawal, R., Imieliński, T., & Swami, A. (1993). Mining association rules between sets of items in large databases. ACM SIGMOD, Volume 22 Issue 2, June 1, 1993, Pages 207-2016. IBM Almaden Research Center, California: San José.
- Agrawal, R., Mannila, H., Srikant, R., Toivonen, H., & Verkamo, A. I. (1996). Fast discovery of association rules. *Advances in Knowledge Discovery and Data Mining*, 12(1), 307-328.
- Aumann, Y., & Lindell, Y. (2003). A statistical theory for quantitative association rules. *Journal of Intelligent Information Systems*, 20(3), 255-283.
- Basu, C., Hirsh, H., & Cohen, W. (1998, July). Recommendation as classification: Using social and content-based information in recommendation. In *Aaai/iaai* (pp. 714-720).
- Berry, M. J., & Linoff, G. S. (2004). *Data mining techniques: for marketing, sales, and customer relationship management*. Indianapolis, Indiana. Wiley Publishing, Inc.
- Blattberg, R., Pyöng-do, K. & Neslin, S. (2008). *Database Marketing : Analyzing and Managing Customers*. Springer.
- Bloemer, J., & De Ruyter, K. (1998). On the relationship between store image, store satisfaction and store loyalty. *European Journal of marketing*, 32(5/6), 499-513.
- Borja, R. P. (2009). *Merchandising. Teoría, práctica y estrategia*. ESIC Editorial.
- Boyd, D., & Ellison, N. (2010). Social network sites: definition, history, and scholarship. *IEEE Engineering Management Review*, 3(38), 16-31.
- Boyd, B. K., Takacs Haynes, K., Hitt, M. A., Bergh, D. D., & Ketchen Jr, D. J. (2012). Contingency hypotheses in strategic management research: Use, disuse, or misuse?. *Journal of Management*, 38(1), 278-313.
- Bramer, M. (2007). *Principles of data mining* (Vol. 180). London: Springer.
- Burke, R. (2002). Hybrid recommender systems: Survey and experiments. *User modeling and user-adapted interaction*, 12(4), 331-370.
- Burke, R. (2007). Hybrid web recommender systems. In *The adaptive web* (pp. 377-408). Springer, Berlin, Heidelberg.
- Chiang, R. H., Goes, P., & Stohr, E. A. (2012). Business intelligence and analytics education, and program development: A unique opportunity for the information systems discipline. *ACM Transactions on Management Information Systems (TMIS)*, 3(3), 12.
- Chiu, S., & Tavella, D. (2008). *Data mining and market intelligence for optimal marketing returns*. Routledge.
- Davidson, J., Liebold, B., Liu, J., Nandy, P., Van Vleet, T., Gargi, U. & Sampath, D. (2010). The YouTube video recommendation system. In *Proceedings of the fourth ACM conference on Recommender systems* (pp. 293-296). ACM.
- Díez, E., Landa, J., & Navarro, G. A. (1996). *Merchandising: Teoría y práctica*. Pirámide, Madrid, 139.
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI magazine*, 17(3), 37.
- Goh, D. H., & Ang, R. P. (2007). An introduction to association rule mining: An application in counseling and help-seeking behavior of adolescents. *Behavior Research Methods*, 39(2), 259-266.
- González, E. V., & Broitman, I. A. E. (2012). *Minería de datos*. Consultada en internet el día, 25.
- Good, N., Schafer, J. B., Konstan, J. A., Borchers, A., Sarwar, B., Herlocker, J., & Riedl, J. (1999). Combining collaborative filtering with personal agents for better recommendations. *AAAI/IAAI*, 439.
- Gundecha, P., & Liu, H. (2012). Mining social media: a brief introduction. *Tutorials in Operations Research*, 1(4), 1-17.
- Hahsler, M., Hornik, K., & Reutterer, T. (2006). Implications of probabilistic data modeling for mining association rules. In *From Data and Information Analysis to Knowledge Engineering* (pp. 598-605). Springer, Berlin, Heidelberg.
- Hand, D. J. (2007). *Principles of data mining*. Drug safety, 30(7), 621-622.
- Herschel, R. T., & Jones, N. E. (2005). Knowledge management and business intelligence: the importance of integration. *Journal of Knowledge Management*, 9(4), 45-55.
- Hüsemann, B., Lechtenbörger, J., & Vossen, G. (2000). Conceptual data warehouse modeling. In *Proc. of the 2nd International Workshop on Design and Management of Data Warehouses*, 6 -11.
- Kamakura, W. A., Kim, B. D., & Lee, J. (1996). Modeling preference and structural heterogeneity in consumer choice. *Marketing Science*, 15(2), 152-172.
- Kannan, P. K., & Wright, G. P. (1991). Modeling and testing structured markets: A nested logit approach. *Marketing Science*, 10(1), 58-82.

- Kantardzic, M. (2011). *Data mining: concepts, models, methods, and algorithms*. Hoboken, New Jersey. John Wiley & Sons, Inc.
- Kim, H. K., Kim, J. K., & Chen, Q. Y. (2012). A product network analysis for extending the market basket analysis. *Expert Systems with Applications*, 39(8), 7403-7410.
- Kotsiantis, S., & Kanellopoulos, D. (2006). Association rules mining: A recent overview. *GESTS International Transactions on Computer Science and Engineering*, 32(1), 71-82.
- Krishnamurthi, L., & Raj, S. P. (1988). A model of brand choice and purchase quantity price sensitivities. *Marketing Science*, 7(1), 1-20.
- Liu, S., Wang, L., & Huang, W. W. (2017). Effects of process and outcome controls on business process outsourcing performance: Moderating roles of vendor and client capability risks. *European Journal of Operational Research*, 260(3), 1115-1128.
- Negash, S. (2004). Business intelligence. *Communications of the association for information systems*, 13(1), 15.
- Nonaka, I., & Takeuchi, H. (1995). *The knowledge-creating company: How Japanese companies create the dynamics of innovation*. Oxford university press.
- Nzivo, C. N., & Chuanfu, C. (2013). International students' perception of library services and information resources in Chinese academic libraries. *The Journal of Academic Librarianship*, 39(2), 129-137.
- Martínez, B. (2016). Escaparatismo y diseño de espacios comerciales. Ediciones Paraninfo, SA.
- Marwick, A. D. (2001). Knowledge management technology. *IBM systems journal*, 40(4), 814-830.
- McAuley, J., Pandey, R., & Leskovec, J. (2015). Inferring networks of substitutable and complementary products. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Pages. 785-794).
- Musalem, A., Aburto, L., & Bosch, M. (2018). Market basket analysis insights to support category management. *European Journal of Marketing*.
- Paiva, E. L., & Gonçalves, C. R. (2008). Organisational knowledge and industry dynamism: an empirical analysis. *International Journal of Innovation and Learning*, 5(1), 66-80.
- Park, M., & Lennon, S. J. (2009). Brand name and promotion in online shopping contexts. *Journal of Fashion Marketing and Management: An International Journal*, 13(2), 149-160.
- Pazzani, M. J. (1999). A framework for collaborative, content-based and demographic filtering. *Artificial intelligence review*, 13(5-6), 393-408.
- Pazzani, M. J., & Billsus, D. (2007). Content-based recommendation systems. In *The adaptive web* (pp. 325-341). Springer, Berlin, Heidelberg.
- Porcel, C., Moreno, J. M., & Herrera-Viedma, E. (2009). A multi-disciplinar recommender system to advice research resources in university digital libraries. *Expert Systems with Applications*, 36(10), 12520-12528.
- Raeder, T., & Chawla, N. V. (2009). Modeling a Store's Product Space as a Social Network. In *2009 International conference on advances in social network analysis and mining* (pp. 164-169). IEEE.
- Raorane, A. A., Kulkarni, R. V., & Jitkar, B. D. (2012). Association rule-extracting knowledge using market basket analysis. *Research Journal of Recent Sciences*
- Rathnavell, J., & Kelkar, K. (2015). Personalized Book Recommendation System. *International Journal Of Engineering And Computer Science*, 21149-21153.
- Rompré, L., Biskri, I., & Meunier, J. G. (2017). Using association rules mining for retrieving genre-specific music files. In *The Thirtieth International Flairs Conference*.
- Roldán, J. L., Carrión, G. A. C., & González, J. L. G. (2012). Los sistemas de inteligencia de negocio como soporte a los procesos de toma de decisiones en las organizaciones. *Papeles de economía española*, (132), 239-260.
- Savinov, A. (2004). *Mining dependence rules by finding largest itemset support quota*. In *Proceedings of the 2004 ACM symposium on Applied computing* (Pages 525-529).
- Segarra Roca, P. (2007). *Influencia de la heterogeneidad del mercado en la intención de comportamiento del consumidor: Respuestas a la actividad relacional en la distribución de gran consumo* (Tesis Doctoral, Universitat Rovira i Virgili).
- Silverstein, C., Brin, S., & Motwani, R. (1998). Beyond market baskets: Generalizing association rules to dependence rules. *Data Mining and Knowledge Discovery*, 2(1), 39-68.
- Srikant, R., & Agrawal, R. (1996). *Mining quantitative association rules in large relational tables*. *ACM SIGMOD*, Volume 25 Issue 2, Pages 1-12).
- Tang, K., Chen, Y. L., & Hu, H. W. (2008). Context-based market basket analysis in a multiple-store environment. *Decision Support Systems*, 45(1), 150-163.
- Videla-Cavieles, I. F., & Rios, S. A. (2014). Extending market basket analysis with graph mining techniques: A real case. *Expert Systems with Applications*, 41(4), 1928-1936.
- Wang, H., & Wang, S. (2008). A knowledge management approach to data mining process for business intelligence. *Industrial Management & Data Systems*, 108(5), 622-634.
- Wellhoff, Alain & Masson, Jean-Emile (2005). *El Merchandising: bases, nuevas técnicas*,

gestión de categorías. Deusto S.A. ediciones.

Wu, C. H., Lee, T. Z., & Kao, S. C. (2004). Knowledge discovery applied to material acquisitions for libraries. *Information processing & management*, 40(4), 709-725

Yi, K., Chen, T., & Cong, G. (2018). Library personalized recommendation service method based on improved association rules. *Library Hi Tech*, 36(3), 443-457.

Zhang, C., & Zhang, S. (2002). *Association rule mining: models and algorithms*. Springer-Verlag.

**ANEXO 1: PROTOTIPO DE APLICACIÓN MÓVIL PARA EL SERVICIO DE RECOMENDACIÓN DE LA BIBLIOTECA DE LA UNIVERSIDAD DE SEVILLA**

