



Universidad  
Internacional  
de Andalucía

## TÍTULO

**DESARROLLO DE UN MODELO DE SISTEMA DE RECOMENDACIÓN  
DE VENTAS PARA LA TOMA DE DECISIONES EN EMPRESAS –  
APLICACIÓN DE UN CASO REAL**

## AUTORA

**Ana Isabel Marín García**

Tutor	<b>Esta edición electrónica ha sido realizada en 2024</b>
Institución	Dr. D. Gonzalo Antonio Aranda Corral
Curso	Universidad Internacional de Andalucía
©	<i>Máster de Formación Permanente en Big Data (2022/23)</i>
©	Ana Isabel Marín García
Fecha documento	De esta edición: Universidad Internacional de Andalucía 2023



Universidad  
Internacional  
de Andalucía



**Atribución-NoComercial-SinDerivadas  
4.0 Internacional (CC BY-NC-ND 4.0)**

Para más información:

<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.es>

<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.en>

Desarrollo de un modelo de sistema de recomendación de ventas para la toma de decisiones en empresas - Aplicación de un caso real

Por:

Alumna: Ana Isabel Marín-García

Tutor: Gonzalo A. Aranda Corral

Una tesis presentada de conformidad con los requisitos  
para el Máster en Big Data

Universidad Internacional de Andalucía



2023



Developing a Sales Recommendation System Model and decision making in  
Companies - A Real Case Application

By:

Alumn: Ana Isabel Marín-García

Supervisor: Gonzalo A. Aranda Corral

A thesis submitted in accordance with the requirements  
for the Master in Big Data

International University of Andalusia



2023



## Resumen

Los sistemas de recomendación se utilizan con cada vez más frecuencia en las empresas para proporcionar sugerencias personalizadas y automáticas a sus clientes sobre nuevos productos y servicios a contratar. Son herramientas que utiliza algoritmos de aprendizaje automático para sugerir productos, servicios o contenido que puedan ser de interés para un usuario en particular.

En este proyecto se desarrolla un sistema de recomendación de tipo colaborativo a partir del análisis formal de conceptos (FCA). Se trata de una técnica de minería de datos que se utiliza para extraer información y conocimiento de conjuntos de datos. FCA es especialmente útil para analizar conjuntos de datos relacionales, en los que cada registro tiene un conjunto de atributos y valores asociados. Además, permite la identificación de patrones y la agrupación de datos en conceptos formales, lo que facilita la interpretación y el análisis de los datos.

El objetivo de nuestro sistema de recomendación es mejorar la experiencia del usuario al personalizar las sugerencias que recibe, y poder incorporarlas en los sistemas informáticos de las tiendas de la empresa.

Se utilizan datos reales de una empresa de muebles que cuenta con nueve sedes a lo largo de Andalucía, con una facturación anual alrededor de 600 mil € anuales (2020) y que se encuentra en proceso de crecimiento.

En resumen, este proyecto tiene como objetivo desarrollar un sistema de recomendación eficiente y personalizado utilizando técnicas de minería de datos y aprendizaje automático, con el fin de poder incorporar estas recomendaciones de productos a comprar por nuestros clientes en los sistemas informáticos de las tiendas de la empresa.

Los resultados del SR concluyen que se ha logrado proporcionar una recomendación de ventas adecuada y personalizada para nuestros clientes basándonos en los datos históricos de clientes similares.

**Palabras clave:** Sistema de recomendación, Análisis formal de conceptos, FCA, Minería de datos, Aprendizaje automático, Filtros Colaborativos, Reglas de asociación, Empresa, Datos Reales.





## Abstract

Recommender systems are increasingly used by companies to provide personalized and automatic suggestions to their customers about new products and services to hire. These tools use machine learning algorithms to suggest products, services or content that may be of interest to a particular user.

In this Project, a collaborative filtered recommendation system is developed with real data, based on the formal analysis of concepts (FCA). It is a data mining technique used to extract information and knowledge from data sets. FCA is especially useful for analyzing relational data sets, where each record has a set of associated attributes and values. In addition, it allows the identification of patterns and the grouping of data into formal concepts, which facilitates the interpretation and analysis of the data.

The objective of our recommendation system is to improve the user experience by personalizing the suggestions you receive, and being able to incorporate them into the computer systems of the company's stores.

Real data from a furniture company that has nine offices throughout Andalusia, with an annual turnover of around €600,000 per year (2020) and which is in the process of growth, is used.

In summary, this project aims to develop an efficient and personalized recommendation system using data mining and machine learning techniques, in order to be able to incorporate these recommendations of products to buy by our customers in the computer systems of the stores of the business.

Results conclude that it has been possible to provide an adequate and personalized sales recommendation for our clients based on the historical data of similar clients.

**Key words:** Recommender System, Formal Concept Analysis, FCA, Data Mining, Machine Learning, Collaborative Filters, Association Rules, Enterprise, Real Data.



**Agradecimientos**

Quisiera dedicar este proyecto a mi familia, por estar siempre conmigo y apoyarme a lo largo de mi vida.

También deseo agradecer a mi tutor el Dr. Gonzalo Aranda por su gran apoyo, orientación y dedicación durante la realización de esta tesis. Gracias por su disponibilidad siempre que lo necesité y darme la oportunidad de ser parte de este proyecto. He disfrutado investigando sobre los sistemas de recomendación, un tema que realmente me apasiona.

Por último, me gustaría dedicar esta tesis a Álvaro, este proyecto no hubiera sido posible sin el cariño y apoyo que he recibido de tu parte.

**Acknowledgments**

I would like to dedicate this project to my family, for always staying with me and supporting me throughout my life.

I also wish to thank my supervisor Dr. Gonzalo Aranda for his great support, guidance and dedication during this thesis. Thank you for your availability whenever I needed it and give me the opportunity to be part of this project. I have enjoy researching about the topics which I am truly passionate.

Finally, I would like to dedicate this thesis to Álvaro, this project would not have been possible without the love and support I received from you.



**Índice de Contenidos**

1	Introducción .....	1
1.1	Breve historia y motivación .....	1
1.2	Identificación del problema .....	1
1.3	Solución al problema .....	2
1.4	Estructura y metodología aplicada.....	2
1.5	Objetivos del proyecto .....	3
2	Estado del arte: Revisión de la literatura.....	5
2.1	Selección de publicaciones relevantes .....	7
3	Marco teórico: Conceptos básicos .....	9
3.1	Análisis Formal de Conceptos (FCA).....	9
3.2	Sistemas de recomendación .....	13
4	Ciclo de ciencia de datos.....	15
4.1	Recopilación de datos .....	15
4.2	Preprocesamiento y almacenamiento de datos .....	16
4.3	Entrenamiento y aplicación del modelo ML basado en FCA .....	24
4.4	Validación de datos .....	31
4.5	Visualización de los resultados .....	33
5	Resultados y conclusión.....	35
6	Futuros proyectos y líneas de investigación.....	37
	Bibliografía .....	39
	Selección artículos importantes .....	41
	<b>Anexos</b> .....	43



## Índice de Tablas

Tabla 1: FCA y SR metodología de búsqueda. Elaboración Propia.....	6
Tabla 2: Análisis en detalle de las publicaciones más relevantes de Scopus. Elaboración Propia.....	7
Tabla 3: Crecimiento anual según el volumen de Facturas emitidas. Elaboración Propia.....	18
Tabla 4 : Calidad de resultados en las recomendaciones. Elaboración Propia.....	31
Tabla 5: Resumen resultados del sistema de Recomendación. Elaboración Propia.....	33

## Índice de Ilustraciones

Ilustración 1: Contexto formal (Conexp). Elaboración Propia.....	9
Ilustración 2: Implicaciones (Conexp). Elaboración Propia.....	11
Ilustración 3: Reglas de asociación (Conexp). Elaboración Propia.....	11
Ilustración 4: Retículo (Conexp). Elaboración Propia.....	12
Ilustración 5: Tabla principal del modelo de datos. Elaboración Propia.....	16
Ilustración 6: Modelo de datos relacional. Elaboración Propia.....	17
Ilustración 7: Importación de datos de la BBDD con Microsoft Query. Elaboración Propia.....	18
Ilustración 8: Ciclo Business Intelligence de Microsoft. Fuente: SAS Website.....	23
Ilustración 9: Distribución Clientes con Power BI y Power Maps. Elaboración Propia.....	23
Ilustración 10: Matriz binaria, como imput al FCA. Elaboración Propia.....	24
Ilustración 11: Matriz de Entrada en Concep Explorer del FCA. Elaboración Propia.....	25
Ilustración 12: Reglas de Asociación en Concep Explorer. Elaboración Propia.....	26

## Índice de Gráficas

Gráfica 1: Volumen anual de publicaciones en Scopus. Elaboración Propia.....	5
Gráfica 2: Distribución de ventas mensual. Elaboración Propia.....	19
Gráfica 3: Distribución de Ingresos mensual. Elaboración Propia.....	19
Gráfica 4: Distribución horaria de ventas. Elaboración Propia.....	20
Gráfica 5: Ranking Vendedores. Elaboración Propia.....	20
Gráfica 6: Calidad de Servicio medido según días de Delivery. Elaboración Propia.....	21
Gráfica 7: Esquema conceptual: modelo de generación de recomendaciones. Elaboración Propia.....	27
Gráfica 8: Esquema de relación de los antecedentes existentes con el perfil de histórico de compras o, en su defecto, con el genérico de los consecuentes. Elaboración Propia.....	29





# 1 Introducción

En general, las empresas tienen como objetivo vender la mayor cantidad posible de productos y servicios al precio más adecuado para obtener el mayor margen de beneficio posible, manteniendo una buena relación e imagen de empresa con sus clientes. Además, es importante ofrecer un servicio postventa de alta calidad para conseguir que tus clientes vuelvan y amplíen sus compras.

En este proyecto se realiza un Sistema de Recomendaciones (SR) de ventas basado en Análisis Formal de Conceptos (FCA) para mejorar el proceso comercial de la empresa a estudiar, con el principal objetivo de mejorar el servicio prestado a nuestros clientes al ofrecerles mejores sugerencias de compras añadidas que pueden adquirir con sus compras.

## 1.1 Breve historia y motivación

El aprendizaje automático (en adelante también ML) es una rama de la inteligencia artificial que estudia algoritmos capaces de aprender de forma autónoma, directamente a partir de datos de experiencia. Durante la última década, las técnicas de aprendizaje automático se han vuelto más importantes, sin embargo, las aplicaciones industriales y la toma de decisiones basadas en datos aún se limitan a un pequeño grupo de grandes empresas.

## 1.2 Identificación del problema

Este proyecto surge ante la solicitud de la empresa de colchones de mejorar las recomendaciones que dan sus empleados a los clientes sobre qué nuevos productos adquirir o cuáles son aquellos que mejor les viene según sus necesidades.

Actualmente en las empresas, a menudo los empleados recomiendan productos a sus clientes basados en modelos mentales subjetivos, reflejados únicamente por su experiencia o gustos personales. Se suelen tomar decisiones o recomendaciones mediante prueba y error, y también según el tipo de cliente que consideren que es. Los sistemas de recomendación optimizan estas tareas y se ocupan de recomendar un producto o asignar una calificación a un artículo, basándose en históricos de compras de esos clientes o de clientes similares a ellos, mejorando así las tasas de acierto. Así, obtendremos una lista ordenada de productos de más a menos según el grado de interés que puedan tener nuestros clientes.

Por ello, el problema que resolver en este proyecto es identificar productos en los que nuestros clientes puedan estar más interesados para recomendárselo a nuestros clientes, bajándonos en sus históricos de compras o en el de clientes similares a ellos.

Un ejemplo de empresa de éxito que utiliza sistemas de recomendación de productos para sus clientes es Amazon, quien atribuye el éxito del 35% de su facturación, alrededor de cuatrocientos mil millones de dólares en 2020.

### 1.3 Solución al problema

Con este problema identificado, nos enfocamos en este proyecto en desarrollar un modelo de sistema de recomendación de ventas para apoyar la toma de decisiones de los empleados de la empresa de colchones. Así, se busca mejorar la satisfacción de los clientes y la atención personalizada de los empleados en las tiendas, al poder indicarles de manera rápida y automática qué productos recomendar a los clientes.

Se ha optado por desarrollar el modelo de recomendación sobre el software de ventas ya existentes en las tiendas físicas, incorporando esta nueva funcionalidad sobre los programas y dispositivos tablets de los vendedores. Además, este modelo debe de ser fácilmente replicado en todas las tiendas físicas de Andalucía, e incluso en otras empresas similares.

En una segunda fase se propondrá también adaptar este sistema de recomendación de manera sencilla y rápida al ámbito del telemarketing, pudiendo ofrecerles a nuestros clientes que lo hayan solicitado y aceptado previamente ofertas postventas. De momento, esto queda fuera del alcance de este proyecto.

Con este alcance definido, se desarrolla todo un proceso de ciclo de ciencia de datos, desde la obtención de los datos reales aportados por la empresa, pasando por el análisis y detalle de los resultados y acabando por la visualización y toma de decisiones a partir de las recomendaciones de los productos y servicios obtenidos.

Debido al gran volumen de datos de ventas históricas que tenemos de las ventas, comenzamos identificando grupos de clientes y sus necesidades para captar qué producto los clientes consideran importante, para así agruparlos y trabajar con diferentes agrupaciones de clientes. Los datos de entrada será el histórico de factura del que ya disponemos de nuestros clientes.

### 1.4 Estructura y metodología aplicada

Este proyecto está estructurado de la siguiente manera.

En el primer bloque, una vez introducido el tema, el problema y la solución a desarrollar el proyecto en los apartados anteriores, se exponen los objetivos y alcance del proyecto y la metodología a seguir.

A continuación, el segundo bloque expone el estado de arte y revisión de la literatura actual existente sobre casos reales que apliquen FCA para Sistemas de Recomendación.

El tercer bloque introduce los fundamentos teóricos de los conceptos aplicados en el proyecto, como la teoría del aprendizaje automático, el Análisis formal de conceptos (FCA) y los Sistemas de Recomendación (SR).

El cuarto bloque es el más importante y extenso. Desarrolla la parte práctica exponiendo los diferentes pasos aplicados para realizar el Sistema de Recomendación. Se sigue la metodología del Ciclo de Ciencia de los datos; desde la recopilación de los datos hasta la visualización de los resultados.

El quinto bloque expone en detalle los resultados y aporta una breve reflexión sobre las recomendaciones obtenidas. Además, expone las conclusiones alcanzadas.

Por último, el sexto bloque detalla posibles futuras líneas de investigación o aplicaciones prácticas de mejora y optimización de los datos de la empresa.

## 1.5 Objetivos del proyecto

Los principales objetivos que buscamos en este proyecto son los siguientes

1. Desarrollar un modelo de recomendación de ventas, basándonos en los datos históricos de clientes similares, para incorporar en los programas de nuestras tiendas.
2. Introducción y aplicación de los diferentes conceptos de aprendizaje automático necesarios.
3. Revisión extensa de la literatura FCA de sistemas de recomendación con las características del esquema de clasificación a tener en cuenta. Selección de los artículos más relevantes para nuestro enfoque.
4. Descripción del problema del caso real de estudio y cómo resolverlo.
5. Identificación de objetivos y restricciones que necesitan ser modelados con Inteligencia Artificial a través del desarrollo del Data Science Cycle.
6. Comprensión y descripción el modelo matemático, el FCA y el Código creado para el Sistema de Recomendación comercial.
7. Explicación de los ejemplos reales de recomendaciones que muestra nuestro modelo.



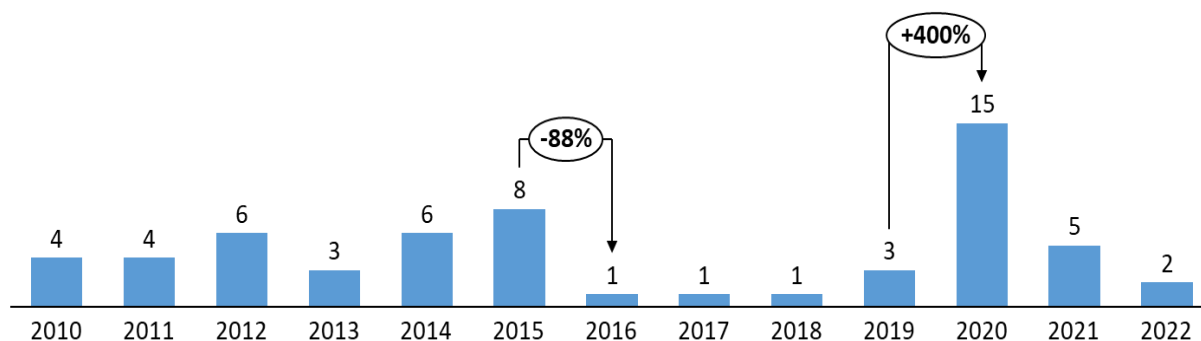
## 2 Estado del arte: Revisión de la literatura

En este bloque se realiza un estudio inicial de la literatura existente y de los últimos años sobre sistemas de recomendación. El principal objetivo es identificar publicaciones que muestren características parecidas a nuestro proyecto en el que apoyarnos y poder así aportar conocimientos a este estudio.

En primer lugar, se desarrolla la metodología y se seleccionan los documentos a incluir en el estudio antes de analizarlos descriptivamente. Las palabras claves por los que se filtran estos documentos en Scopus son aquellos que contengan, en inglés, en el título, abstract o keywords los siguientes términos: "Recommendation System", "Recommender Systems", "FCA", o "Formal concept Analysis". Estos criterios se indican en la Tabla 1.

Una vez definidas estas palabras claves, se utilizan para buscar en la base de datos de Scopus, a fecha agosto de 2022. Se limita al tipo de documento "artículos" y "documentos de conferencias". El rango de fechas de los artículos publicados es de 2010 a 2022.

Con estos primeros criterios de búsqueda obtenemos los primeros 59 resultados. Como se indica en la Gráfica 1, vemos que el 52,3% de los resultados fueron publicados entre 2010 y 2015 y que entre 2016 y 2019 únicamente se realizaron 6 publicaciones. En 2020, correspondiendo con el año del Covid-19, fue el año con más éxito de publicaciones sobre FCA y RS, con 15 publicaciones, correspondientes al 25,4% del total.



Gráfica 1: Volumen anual de publicaciones en Scopus. Elaboración Propia

El área de conocimiento más destacado ha sido el de Computer Science, seguido de Engineering y Mathematics. Por otra parte, China es el país que más artículos publica con esta temática, seguido de España.

En una segunda búsqueda en Scopus, añadimos un nuevo filtro para identificar que se traten casos reales de aplicación en empresas o con datasets reales. Con esto, pasamos de obtener 59 publicaciones a 15.

A continuación, la búsqueda en la base de datos se complementó con una búsqueda de bola de nieve hacia adelante y hacia atrás para eliminar artículos irrelevantes. Se partió de los 59 resultados de la primera búsqueda en Scopus y de descartaron los artículos que no estaban escritos en inglés o español, por ejemplo, en chino o portugués.

<b>Tipo de filtro</b>	<b>Descripción y directrices</b>	<b>Resultado</b>
<b>Criterios incluidos</b>	Documentos: Artículos identificados en la búsqueda en la base de datos Scopus. Solo artículos publicados y artículos de conferencias con un enfoque en el Sistema de Recomendación de análisis formal de conceptos Idioma: Limitado al inglés Duración: 2010 a 2022 (enero a julio) Tipo de documento: Artículos, Documentos de conferencias, Reviews	
<b>Palabras claves buscadas</b>	<i>(TITLE-ABS-KEY (formal AND concept AND analysis ) AND TITLE-ABS-KEY ( FCA ) AND TITLE-ABS-KEY ( recommendation AND system ) AND TITLE-ABS-KEY ( Recommender AND Systems ) ) AND PUBYEAR &gt; 2009 AND ( LIMIT-TO ( PUBSTAGE , "final" ) )</i>	59
<b>Enfoque</b>	Se consolidaron los resultados de Scopus, que no aplican, se eliminaron y se seleccionaron las publicaciones relevantes e incluyendo, mediante la lectura de los resúmenes	-14
<b>Comprobación</b>	Se analizaron todos los Abstract de los trabajos identificados y se examinó su contenido. Algunos de ellos excluyeron:	-4
<b>Muestra final</b>		<b>41</b>

Tabla 1: FCA y SR metodología de búsqueda. Elaboración Propia

Además, todas las obras preseleccionadas fueron leídas de forma independiente y estudiadas en profundidad para comprobar su contenido, en la mayoría a partir del abstract, recurriendo al resto del artículo en los casos oportunos. Después de este análisis, se encontró que algunos de los documentos no eran relevantes y, por lo tanto, se excluyeron dando finalmente el tamaño final de la muestra.

Para concluir, el siguiente apartado presenta una descripción y clasificación general de los contenidos más relevantes sobre los que trata esta muestra final seleccionada para este proyecto.

## 2.1 Selección de publicaciones relevantes

En la Tabla 2 se muestran los que podrían considerarse las publicaciones más importantes encontrados en la revisión de la literatura. Aunque no coinciden perfectamente con todos los sistemas de recomendación que usan funciones de FCA, sí coinciden con la mayoría de los más cercanos.

REFERENCIA	DESARROLLO SR	APLICACIÓN SR	USO FCA	COLABORATIVE FILTER	REAL CASE	FECHA RECIENTE
1. KATARIA S., BATRA U. [2022]	x		x	x		x
2. LIU Z.-H. ET AL. [2022]	x		x	x		x
3. SEFRIQUI BOUJEMAA K. ET AL. [2021]	x		x	x	x	x
4. INDIRA K. ET AL. [2021]	x		x	x	x	x
5. BELLINI E. ET AL. [2020]		x	x			x
6. BUNNELL L. ET AL. [2020]		x	x			x
7. CHEN Y.-H. ET AL. [2020]		x	x			x
8. CORDERO P. ET AL. [2020]	x		x		x	x
9. CHEMMALAR SELVI G. ET AL. [2020]	x		x	x		x
10. G C.S. ET AL. [2020]	x		x	x		x
11. JORRO-ARAGONESES J.L. ET AL. [2020]	x		x	x		x
12. MENŠÍK M. ET AL. [2020]		x	x			x
13. MUANGPRATHUB J. ET AL. [2020]	x		x	x		x
14. VILAKONE P. ET AL. [2020]	x		x	x	x	x
15. CARBONNEL J. ET AL. [2019]	x		x	x	x	x
16. CHEMMALAR SELVI G. ET AL. [2019]	x		x	x	x	x
17. SELVI G.C. ET AL. [2019]	x		x	x		x
18. MEZNI H. ET AL. [2018]	x		x	x	x	x
19. NAİM H. ET AL. [2016]		x	x			
20. AKHMATNUROV M. ET AL. [2015]		x				
21. ALQADAH F. ET AL. [2015]	x				x	
22. BEDEK M.A. ET AL. [2015]		x	x			
23. KASHNITSKY Y. ET AL. [2015]	x		x	x	x	
24. ZHANG W. ET AL. [2015]	x		x	x		
25. ZHANG W. ET AL. [2015]	x		x	x		
26. ZOU C. ET AL. [2015]		x	x			
27. BAU C.-T. ET AL. [2014]	x		x	x	x	
28. CASTELLANOS A. ET AL. [2014]	x		x	x		
29. IGNATOV D.I. ET AL. [2014]	x		x	x	x	
30. IGNATOV D.I. ET AL. [2014]	x		x	x	x	
31. SENATORE S. ET AL. [2013]	x		x	x		
32. BOUAUD J. ET AL. [2012]	x		x	x	x	
33. DE MAIO C. ET AL. [2012]	x		x	x		
34. ENCHEVA S. [2012]	x		x	x		
35. LI X., MURATA T. [2012]	x		x	x	x	
36. XU H.-S., KUANG G.-F. [2012]	x		x	x	x	
37. ZHENG Q., WANG Y. [2012]	x		x	x	x	
38. SUN X., LI J. [2011]		x	x	x	x	
39. LI X., MURATA T. [2010]	x		x	x		
40. LI X., MURATA T. [2010]		x	x	x		
41. CHERNIAHOVSKAYA L. ET AL. [2010]		x	x			

Tabla 2: Análisis en detalle de las publicaciones más relevantes de Scopus. Elaboración Propia

Además, la mayoría de la literatura encontrada trata de casos teóricos de investigación y trabajos académicos, se identifican unos pocos casos prácticos que aplican datos reales de empresas. De esto concluimos que nuestro proyecto va a aportar gran valor a la literatura disponible.

Hemos clasificado los artículos finales seleccionados según las siguientes variables: 1. Desarrollo de un Sistema de recomendación, 2. Aplicación de un Sistema de Recomendación ya existente, 3. Utiliza el FCA 4. Collaborative Filtering, 5. Aplicación de Casos reales, 6. Fecha reciente (últimos 5 años). El detalle con las publicaciones seleccionadas se encuentra al final del documento, en los anexos.





### 3 Marco teórico: Conceptos básicos

#### 3.1 Análisis Formal de Conceptos (FCA)

El Análisis Formal de Conceptos (FCA por sus siglas en inglés de *Formal Concept Analysis*) es una teoría matemática que formaliza la noción de "concepto" y nos permite calcular una jerarquía a partir de una tabla de relaciones. Específicamente, aplicado a los sistemas de etiquetado, hace explícito el conjunto de conceptos implícitos que maneja el usuario en su trabajo de etiquetado. Esta teoría matemática fue introducida por Rudolf Willie y B. Ganter en 1984.

El marco de conceptos proporcionado por FCA representa una estructura intermedia entre el etiquetado (no jerárquico) y las taxonomías clásicas (ontológicas). A partir de datos de entrada tabulares dados, este método encuentra todos los conceptos formales y calcula un retículo de conceptos, en nuestro proyecto finito (no tiene número infinito de elementos) y acotado (tiene máximos o supremo y mínimos o ínfimos).

Dentro de FCA, el contexto formal es la unidad básica de representación del conocimiento. En su libro "Formal Concept Analysis", Ganter, B. y Wille, R. (1999) definen un contexto formal como una terna  $K = (G, M, I)$  donde "G" es un conjunto de objetos, "M" es un conjunto de atributos e "I" es la relación existente entre "G" y "M". En la Ilustración 1 se muestra el Contexto formal utilizado para nuestro modelo en Conexp.

A su vez, postulan que para que un par  $(A, B)$  dentro de un contexto formal  $K = (G, M, I)$  sea un concepto formal verificar que:  $A \subseteq G$ ,  $B \subseteq M$  y que  $A' = B$ ,  $B' = A$ .

Si  $K = (A, B)$  es un concepto formal, dentro de él se pueden identificar dos partes. Por un lado, se encuentra la extensión, formada por el conjunto de objetos "A" que lo componen. Y por el otro lado se puede observar la intensión, formada por el conjunto de atributos "B".

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
155															
217		X													
1509											X				
1835															
2195															
2303											X				
2698											X	X			
3240															
3559											X				
3979															
4940					X									X	
4769															
5555															
5587															
5858															
6672												X			
7279															
8154															
8494															
8713															
9040															
9211		X										X			
9220												X			
9253															
9274															
9444		X									X		X		
9669															
10145															
10329		X									X				
10600															
10897											X				
11014															
11040		X													
11824															
12314												X			
12334															
12477		X													

Ilustración 1: Contexto formal (Conexp). Elaboración Propia

Cuando se extraen los datos, se construye el contexto formal asociado considerando los documentos digitales extraídos como los objetos de nuestro contexto y las etiquetas asociadas a ellos como sus atributos.

Los contextos formales también se pueden definir como un conjunto de implicaciones entre los atributos de este, tales que las implicaciones pueden ser extraídas del contexto y el contexto puede ser reconstruido a partir de las implicaciones.

Las implicaciones son restricciones sobre los atributos, y son de la forma:

$$\{a_i, \dots, a_j\} \rightarrow \{a_k, \dots, a_m\}$$

La idea intuitiva es expresar la relación:

“Todo objeto que satisface los atributos  $\{a_i, \dots, a_j\}$  también satisface los atributos  $\{a_m, \dots, a_k\}$ ”

y formalmente definido como: Implicación entre atributos: Sea  $C = (O, A, I)$  un contexto formal. Una implicación de atributos es un par de subconjuntos de atributos, y normalmente escritos de esta forma  $L \rightarrow R$ . Una implicación  $L \rightarrow R$  es válida en  $C$  si para todo objeto de  $C$  que tiene todos los atributos de  $L$  también tienen todos los atributos de  $R$ . A todas las implicaciones extraídas de un contexto  $C$ .

La unidad básica de razonamiento en FCA es el razonamiento implicacional entre atributos. Este razonamiento se puede ver desde dos puntos de vista. Uno sería el clásico, basado en el cierre implicacional (en la idea de sistema de producción), y el otro sería el razonamiento directo con las implicaciones.

Una vez obtenido las implicaciones, se debe de hablar de la consecuencia lógica y para ello se utilizan las reglas de Armstrong, definidas como: (W. Armstrong, 1974)

$$\frac{}{X \rightarrow Y} \text{ (Identidad)} \quad \frac{X \rightarrow Y}{X \cup Z \rightarrow Y} \text{ (Extensión)} \quad \frac{X \rightarrow Y \quad Y \cup Z \rightarrow W}{X \cup Z \rightarrow W} \text{ (Substitución)}$$

$$\text{con } W, X, Y, Z \subseteq a A.$$

El conjunto de implicaciones de un contexto,  $\text{Imp}(C)$ , puede ser realmente grande, aunque nosotros estamos únicamente interesados en las bases de este conjunto con el menor número de implicaciones posibles. En la Ilustración 2 se muestran las Implicaciones obtenidas en nuestro modelo.

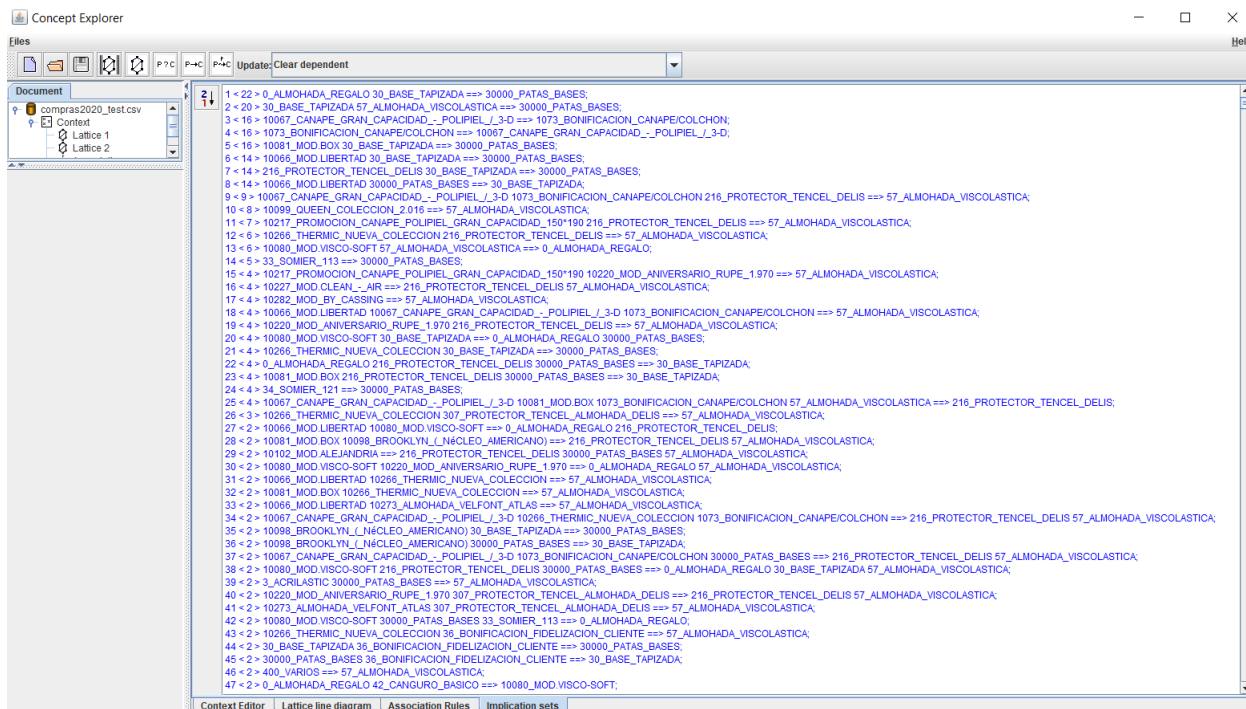


Ilustración 2: Implicaciones (Conexp). Elaboración Propia

El último concepto que se definen las reglas de asociación, que son implicaciones que satisfacen un conjunto de objetos del contexto. Dentro de las reglas de asociación se pueden identificar dos partes, la primera es el soporte, número de objetos que cumplen las reglas, y la segunda es la confianza, que es el porcentaje de objetos que cumplen la premisa y las conclusiones. En la Ilustración 3 se muestran las Reglas de asociación de nuestro modelo en Conexp.

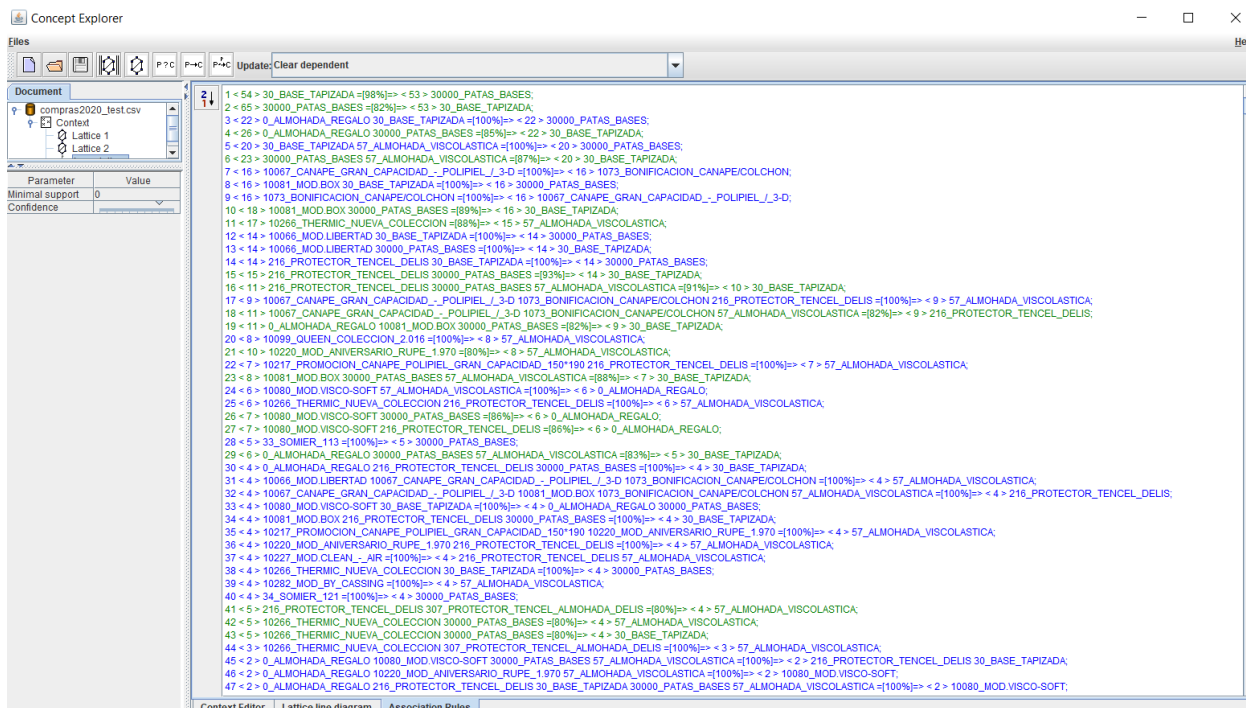


Ilustración 3: Reglas de asociación (Conexp). Elaboración Propia

Las principales ventajas del método FCA son la simplicidad de preparar grandes conjuntos de datos de entrada para el análisis y que el análisis puede resultar con una visión general clara del dominio de entrada y con una lista de atributos importantes trascendencia.

Por último, el retículo de FCA es una representación gráfica de la relación entre conceptos formales y sus atributos en un conjunto de datos. Se utiliza como una herramienta visual para facilitar la interpretación y el análisis de los datos. En la Ilustración 4 se muestra el Retículo obtenido en nuestro modelo.

Se compone a partir de los dos conjuntos de elementos vistos en la tabla de conexp anterior: los conceptos formales y sus atributos. Para construir el retículo de FCA, se dibuja un círculo para cada concepto formal y se colocan los atributos del concepto en el interior del círculo. Los objetos que pertenecen al concepto se representan mediante una línea que conecta el círculo del concepto con el objeto correspondiente

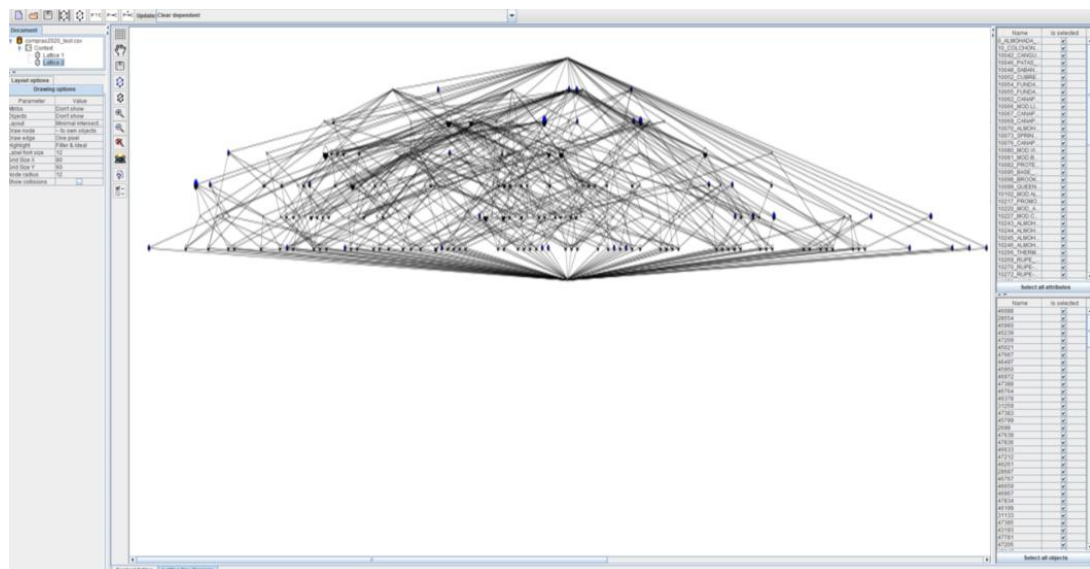


Ilustración 4: Retículo (Conexp). Elaboración Propia

## 3.2 Sistemas de recomendación

Los sistemas de recomendación son sistemas inteligentes que ofrecen al usuario unas sugerencias personalizadas sobre algún elemento en concreto basado en la experiencia. Melville y Sindhvani (2010) en su libro “Recommender Systems: Encyclopedia of Aprendizaje automático” afirman que, “El objetivo de los sistemas de recomendación (SR) es generar recomendaciones válidas de elementos para un conjunto de usuarios a los que le puedan interesar”.

Para llevar a cabo estas recomendaciones, los sistemas de recomendación suelen utilizar la información obtenida del usuario, es decir, sus preferencias y sus comportamientos pasados en relación a los ítems y el propio sistema recomendador.

En función al tipo de información mencionada, se clasifican los tipos de sistemas de recomendación. Para la clasificación, se va a utilizar parte de la expuesta por Charu C Aggarwa (2016), en su libro “Recommender Systems: The Textbook”:

- Sistema de recomendación basado en filtro colaborativo: Basan la recomendación en un grupo de personas que le gusta lo mismo que al sujeto al que se le va a realizar la recomendación.
- Sistema de recomendación basado en contenido: Basan la recomendación en el análisis del contenido de los ítems (descrito por sus características o atributos) y en el perfil de usuario (que realiza un seguimiento de los elementos seleccionados anteriormente).
- Sistema de recomendación híbridos: Basan la recomendación en la combinación de los dos anteriores para mejorar la precisión de la recomendación.

En este trabajo se ha utilizado el sistema de recomendación basado en FCA debido a que los datos con los que se trabaja se amoldan mejor a este tipo de sistemas.



## 4 Ciclo de ciencia de datos

En esta sección se elabora la parte práctica del proyecto; la construcción del sistema de recomendación de filtrado colaborativo, siguiendo los pasos del ciclo de ciencia de los datos.

Data Science es un campo de estudio multidisciplinario que combina habilidades de programación, experiencia en el dominio y conocimiento de estadísticas y matemáticas para extraer información y conocimientos útiles de los datos.

Hay diferentes pasos para desarrollar un ciclo de ciencia de datos correcto que vamos a desarrollar a continuación:

- 4.1 Recopilación de datos
- 4.2 Preprocesamiento y almacenamiento de datos
- 4.3 Extracción del modelo mediante FCA
- 4.4 Validación y creación de nuevos test
- 4.5 Visualización de los resultados

### 4.1 Recopilación de datos

El primer paso es recopilar el conjunto correcto de datos, fundamental para conseguir un modelo de calidad. A veces, esta tarea básica es difícil para algunas organizaciones que ni siquiera saben qué tipo de datos recopilar o dónde residen los datos. Es importante recopilar los datos correctos según el problema comercial que esté tratando de resolver.

Por ejemplo, los datos perfectos para nuestro modelo de sistema de recomendaciones suelen incluir datos de ventas, información de clientes, datos de precios, histórico de productos previamente comprado, patrones de ventas etc.

Existen diferentes mecanismos para recopilar los datos. Las herramientas de recopilación de datos que se han utilizado en este proyecto para extraer datos valiosos de la tienda de muebles es MySQL conectado a la aplicación de facturación de la empresa, que la Universidad había desarrollado algunos años antes para ellos.

Una vez extraído los datos comerciales reales de la empresa (anónima bajo su petición) a partir de su software de facturación, ya podremos comenzar con el primer paso del Data Science Cycle. Nuestra empresa nos ha facilitado datos reales recogidos desde 2017 a 2020. Se trata de cuatro archivos en formato csv. que contienen información de 1. Tabla de datos de sus clientes, 2. Tabla de Facturas, 3. Tabla Tipo de Pago y 4. Tabla de Línea de factura (Indica en cada factura, qué tipo de producto se ha vendido).

El punto de partida es una información referente a registros de venta de productos a clientes en un periodo de tiempo determinado (típicamente facturas de compra, pudiendo contener cada una uno o varios productos).

Nuestra tabla clave o principal de nuestro modelo es aquella que recoge información de las facturas de nuestros clientes y que se muestra en la Ilustración 5.

Id	cdfactura	cdcliente	Fecha_Alta	Hora_Alta	Año_Alta	Mes_Alta	recargo	cdformapago	acuenta	cerrado
5561	31137	28787	03/01/2017 12:05:42		2017		1.0.0	1	20,00	SI
6562	31138	28870	03/01/2017 12:23:22		2017		1.25.0	2	169,12	SI
6563	31139	24125	03/01/2017 13:22:27		2017		1.0.0	1	76,99	SI
6564	31140	4247	03/01/2017 17:51:02		2017		1.0.0	2	16,49	SI
6565	31141	28700	03/01/2017 18:56:37		2017		1.0.0	16	488,99	SI
6566	31142	4324	03/01/2017 19:23:42		2017		1.0.0	2	164,31	SI
6567	31143	28835	04/01/2017 12:03:36		2017		1.0.0	2	230,99	SI
6568	31144	4324	04/01/2017 12:26:13		2017		1.0.0	1	17,99	SI
6569	31145	4247	04/01/2017 12:34:06		2017		1.0.0	2	18,44	SI
6570	31146	28854	04/01/2017 13:14:26		2017		1.0.0	18	50,98	SI
6571	31147	28868	04/01/2017 20:32:29		2017		1.0.0	2	567,49	SI
6572	31148	4380	05/01/2017 12:53:18		2017		1.0.0	2	39,99	SI
6573	31149	4629	05/01/2017 13:07:24		2017		1.0.0	1	59,99	SI
6574	31150	28871	07/01/2017 12:21:52		2017		1.0.0	2	686,92	SI

Ilustración 5: Tabla principal del modelo de datos. Elaboración Propia

A continuación, se incluye el diccionario de datos, mostrando para cada campo incluido en la tabla, su definición, características y el tipo de dato:

- Id: Identificador único generado automáticamente en Access. Código numérico
- Cdfactura: Código identificador de la factura. Código numérico
- Cdcliente: Código identificador del cliente. Código numérico
- Fecha\_Alta: Día, mes y año en que se registra la factura. Código numérico
- Hora\_Alta: Hora, minuto y segundo en la que se registra la factura. Código numérico
- Año\_Alta: Año en el que se registra la Factura. Código numérico
- Mes\_Alta: Mes en el que se registra la Factura. Código numérico
- recargo: Aumento sobre la cuantía del cobro o pago de la factura debido al incumplimiento de la fecha estipulada o partición de la factura. Código numérico
- cdformapago: Código que indica qué tipo de pago se realiza. Código numérico
- acuenta: Cantidad resultante a pagar por el cliente. Código numérico
- cerrado: Si la Factura está finalizada o sigue abierta, requiriendo alguna acción extra. Código texto

## 4.2 Preprocesamiento y almacenamiento de datos

Después de recopilar el conjunto de datos apropiado, se necesita limpiarlos adecuadamente, procesarlos y mantenerlos en una base de datos adecuada. Los datos pueden ser no estructurados, irrelevantes y estar sin filtrar.

Para construir un buen modelo, es necesario resolver valores duplicados y nulos, tipos de datos inconsistentes, datos faltantes, entradas no válidas, formato incorrecto, etc. antes de continuar. El tiempo de la mayoría de los científicos de datos se dedica a la recopilación, limpieza y procesamiento de datos.

Algunos profesionales de datos incluso argumentan que toma el 80% del tiempo dedicado a un proyecto de datos. Si desea crear excelentes modelos de ciencia de datos, debe encontrar y resolver fallas e inconsistencias en el conjunto de datos. Si bien la limpieza de datos es dolorosa y engorrosa, la superará y se beneficiará siempre que permanezca enfocado en el objetivo final.

Para automatizar todo lo posible este paso y optimizar el tiempo de futuras actualizaciones hemos creado nuestra base de datos con la herramienta de Access. De esta forma y mediante Microsoft query y código SQL, se crea una base de datos que integre estos archivos en bruto obtenidos del



software de facturación. Así, podremos ir añadiendo de forma fácil y sencilla nuevas ventas a nuestra base de datos y programar este proceso de carga automática.

En este paso, se aprovecha para homogenizar las variables. Para la variable NIF de cliente, se quitan los puntos, guiones o letras identificadas y se limita a 9 números. Además, se eliminan los repetidos. También se eliminan los productos devueltos, identificados con valores negativos en las facturas.

La tabla base de facturas no contiene qué productos fueron vendidos en cada una, por lo que se hace necesario un primer preprocesado donde se relacione mediante el id de factura con la tabla que contiene un registro por cada producto/s asociado a cada factura.

Adicionalmente, se realiza la relación con las tablas que aportan información sobre el cliente y el medio de pago:

El resultado es una tabla de registros múltiples por factura, donde cada línea representa un cliente y un código y descripción de producto vendido, indicando cuantas unidades se vendieron y a qué precio, pudiendo repetirse en diferentes registros tanto el identificador de cliente como el de producto.

En la Ilustración 6 se muestra el modelo de datos relacional en estrella construido en Power BI, tomando como referencia la tabla de Facturación.

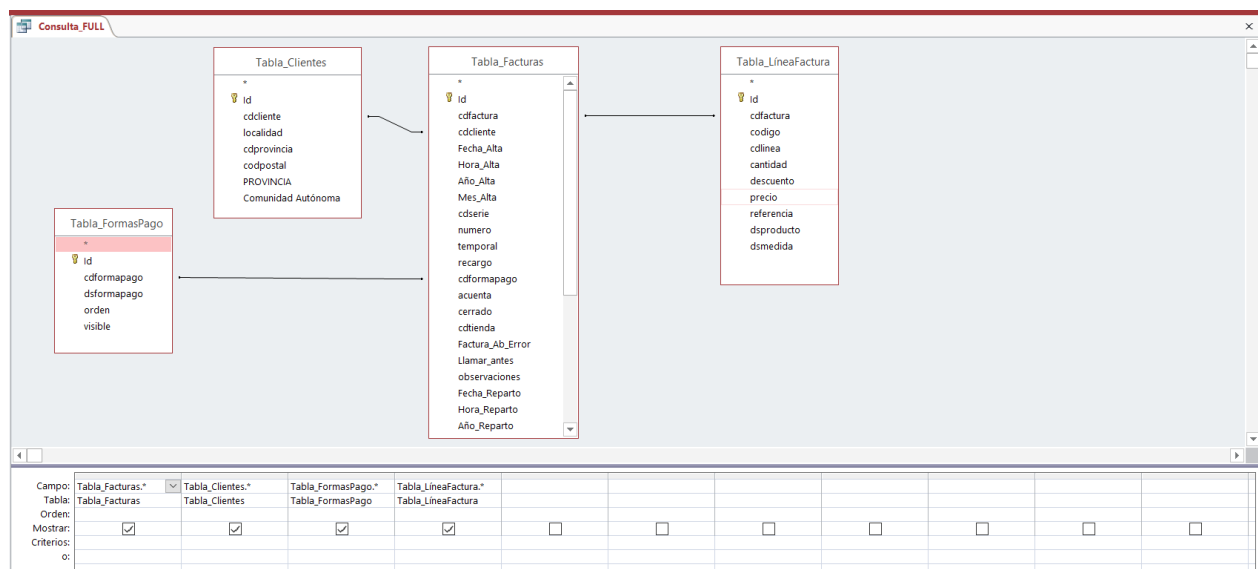


Ilustración 6: Modelo de datos relacional. Elaboración Propia

Una vez que tenemos nuestros datos extraídos, preprocesados, almacenados y fácilmente accesibles, podemos realizar el siguiente paso del Data Science Cycle, el análisis de datos. Para ello, accedemos a los datos de nuestra base de datos en Access mediante Microsoft query y ODBC como muestra la Ilustración 7.

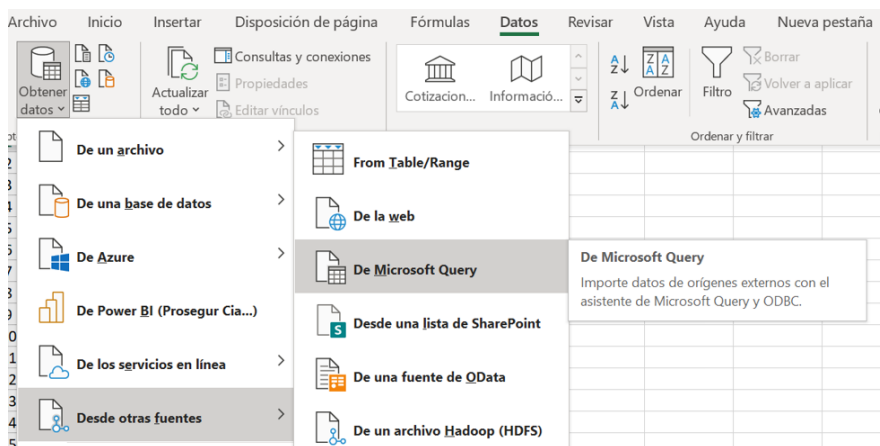


Ilustración 7: Importación de datos de la BBDD con Microsoft Query. Elaboración Propia

El análisis de ventas de las empresas debe centrarse en desarrollar y mejorar una estrategia de crecimiento de ventas a corto, mediano y largo plazo. Para ello, es importante conocer nuestros datos y las características de nuestros clientes, los patrones de compra, los artículos y servicios disponibles y ofrecidos, etc.

Además, el análisis de datos indica qué pasos sobre la limpieza de datos deben realizarse (corrección de valores nulos, eliminación de outliers, etc) y aporta información valiosa para tomar decisiones en torno a la arquitectura adecuada de los modelos de aprendizaje automático en los que servirán de input.

Además, gracias a este apartado, se entiende mejor nuestro tipo de actividad realizada y entender mejor el modelo de recomendación elaborado en capítulos posteriores.

Así, los responsables de la empresa podrán tomar un gran valor sobre las aplicaciones desarrolladas con aprendizaje automático (en este caso, el sistema de recomendación).

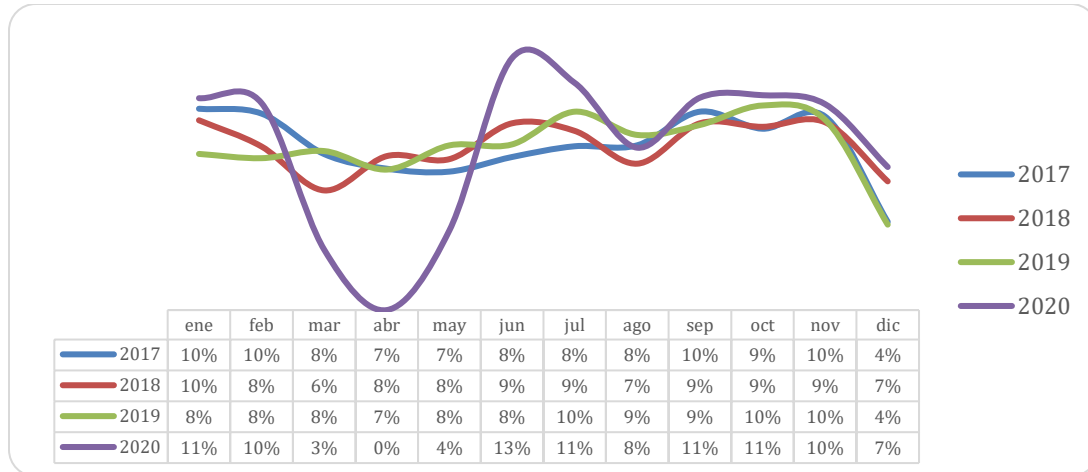
A continuación, se exponen los principales Insights de mejora obtenidos con el análisis cuantitativo de nuestros datos de clientes, facturas, promociones y formas de pago.

En un análisis exploratorio básico, vemos en la Tabla 3, que el nº de facturas se mantiene prácticamente estable durante el periodo temporal disponible. Sin embargo, aumenta la facturación un 10% 2019 frente al año anterior y un 4% 2020 frente al año anterior.

AÑO	#FACTURAS EMITIDAS	VAR VOL FACTURAS	#FACTURACIÓN ANUAL	VAR FACTURACIÓN
2017	1.652	-	560.300€	-
2018	1.616	-2%	524.200€	-6%
2019	1.651	2%	576.100€	10%
2020	1.640	-1%	599.100€	4%
<b>TOTAL</b>	<b>6.559</b>		<b>2.259.700€</b>	

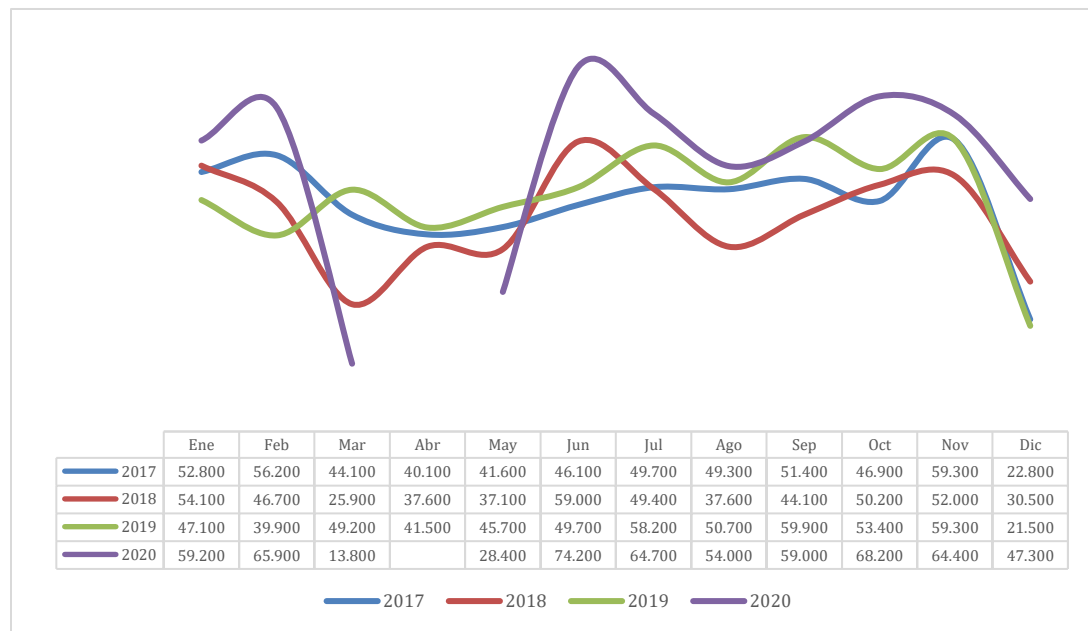
Tabla 3: Crecimiento anual según el volumen de Facturas emitidas. Elaboración Propia

**Distribución de ventas mensual:** Como muestra la Gráfica 2, marzo, abril, mayo y diciembre son los meses de menos facturas emitidas. Por el contrario, octubre y noviembre son los meses de más ventas.



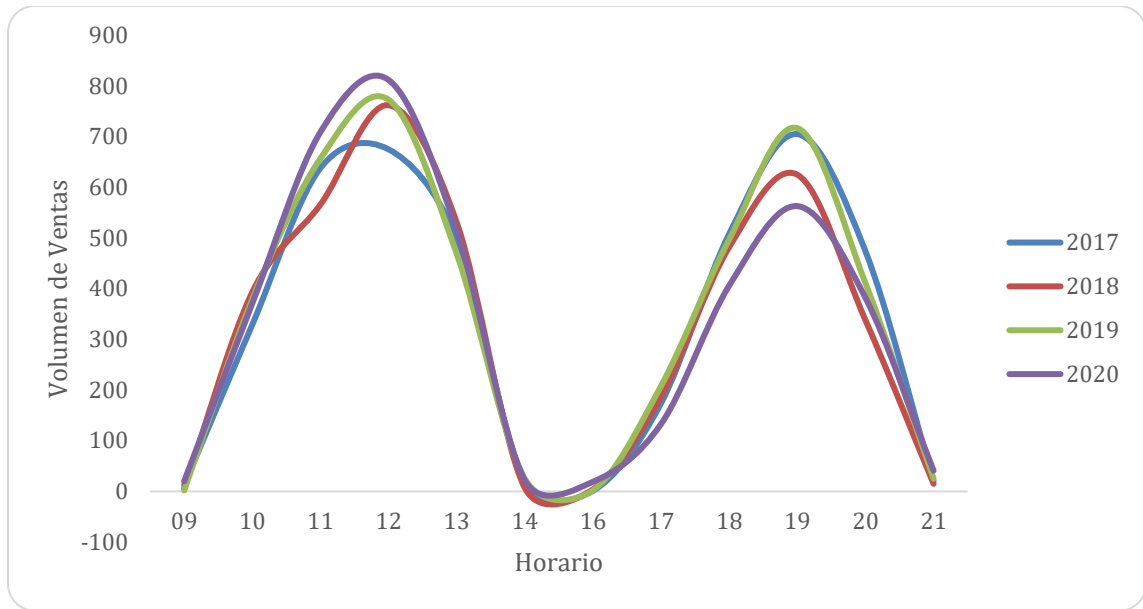
Gráfica 2: Distribución de ventas mensual. Elaboración Propia

**Distribución ingresos mensual:** Vemos en la Gráfica 3 que los meses de menos ventas también coincide con un ingreso menor, al ser productos similares los que se venden cada mes.



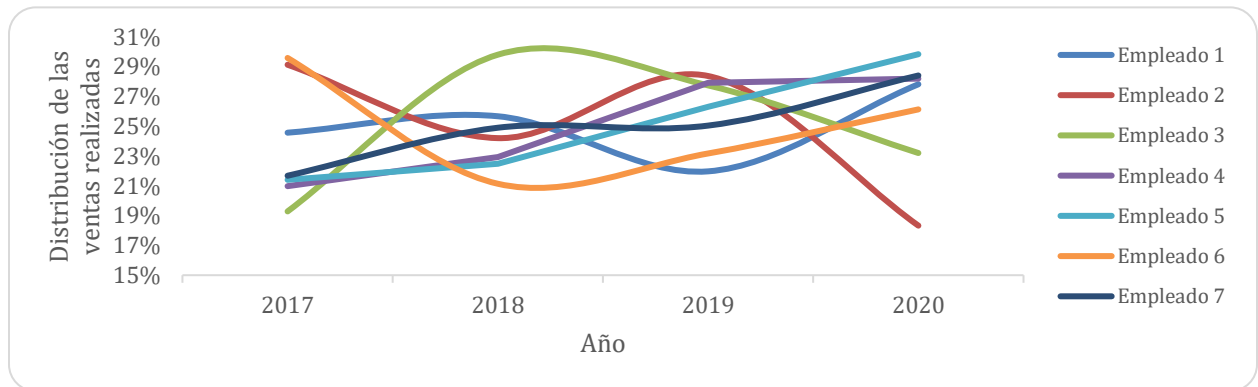
Gráfica 3: Distribución de Ingresos mensual. Elaboración Propia

**Distribución horaria de ventas:** En la Gráfica 4 vemos que se cumple la Ley de Pareto ya que el 85% de las ventas ocurren por la mañana de 11 a 14h y por la tarde de 18 a 21h. Este dato podría ayudar a los departamentos de toma de decisión de la empresa a analizar si abrir una hora más tarde y cerrar una hora antes, ya que, en ese tiempo, se realiza menos del 1% de las ventas totales de la empresa. Además, es curioso ver cómo se produce un cambio en los patrones de compra de los clientes, ya que, en 2020, año de pandemia, aumentan los clientes que compran por la mañana y disminuyen los que compran por la tarde, comparado con años anteriores. Esto puede ser por el fomento del teletrabajo o aumento de paro, donde podríamos indagar.



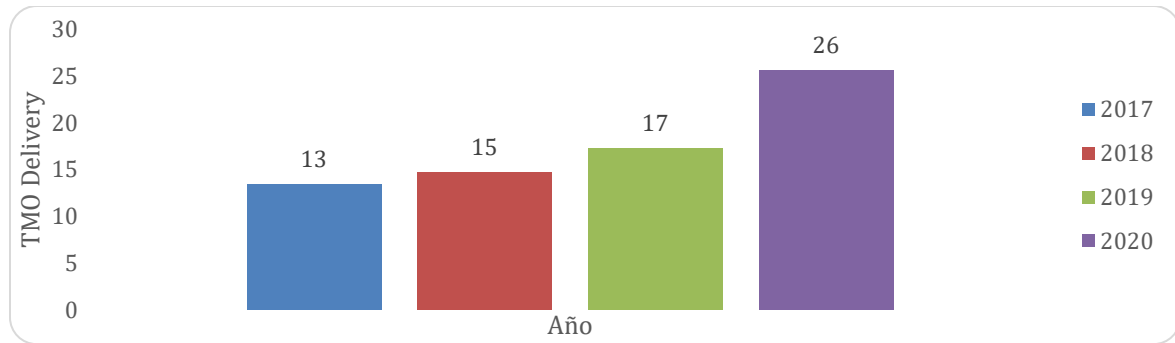
Gráfica 4: Distribución horaria de ventas. Elaboración Propia

**Ranking Vendedores:** En la Gráfica 5 identificamos quienes han sido los mejores vendedores cada año y cómo han ido evolucionando. También se ha dispone del mismo ranking mensual y semanal, permitiendo el seguimiento de los KPI's y productividades de los trabajadores. Los datos de los empleados han sido previamente anonimizados para cumplir con la ley de protección de datos.



Gráfica 5: Ranking Vendedores. Elaboración Propia

**Calidad de Servicio medido según días de espera de entrega:** Como vemos en la Gráfica 6, la calidad de nuestro servicio ha ido disminuyendo cada año, al aumentar el tiempo medio de envío de los productos de 13 días en 2017 a 26 días en 2020. Esto es debido en gran parte a la pandemia y a los grandes problemas de stock asociados.



Gráfica 6: Calidad de Servicio medido según días de Delivery. Elaboración Propia

**Distribución del número de ventas e ingreso según los Productos disponibles:** En cuanto a la dispersión, se vendieron 16.129 unidades de producto correspondiente a 155 referencias distintas de producto, donde los 17 productos más vendidos (el 11%) acumulan el 80% de las unidades vendidas, y a nivel de volumen de ventas el 80% lo consiguen los 14 productos con mayor facturación, dos de ellos que ya no se venden desde 2019 al estar descatalogados.

**Otros Insights obtenidos tras el análisis de datos:** Por último, se han identificado otras áreas sobre las que trabajar estos datos para conseguir mayor rentabilidad de la empresa. Como próximos pasos identificados en futuros proyectos de big data y aprendizaje automático, se destacan algunos:

- Realizar modelos predictivos de ventas.
- Benchmark entre tiendas en función de gastos y ganancias.
- Segmentación de clientes.
- Optimización de Stock: Compras a proveedores: Aumento de eficiencia y reducción de costes y almacén.
- Optimización de rutas de reparto.
- Toma de decisiones de nuevas aperturas por código postal. Para ello habría que hacer estudio de mercado previo.
- Aprendizaje supervisado para identificar qué productos tener en stock, para priorizarlos en la recomendación, aplicarles promociones, o recomendarlos antes: Disminuye los plazos de entrega y espacio en almacén.
- Optimización de órdenes de Compras a proveedores: Aumento de eficiencia y reducción de costes y almacén.

Y como mejoras en el software de facturación disponible actualmente por la empresa, y elaborado por nuestro grupo de investigación, se han identificado las siguientes mejoras:

- Introducir Campo de "Llamar antes de ir"; actualmente aparece en el campo abierto de comentarios cuando el agente se acuerda de codificarlo, no está estandarizado y no es fácilmente explotable.
- Introducir Campo de "Color"; ocurre igual que el caso anterior, actualmente se deja indicado en el campo comentario.
- Introducir Campo de "Retirar lo antiguo (1/0)".

### **Análisis de datos con Power BI:**

Existen diferentes herramientas muy potentes en el mercado para analizar los datos, cada vez más importantes debido a la gran cantidad de información y datos que manejamos. Algunas de las más conocidas y avanzadas son Tableau, Databox, Datawrapper, Zoho Analytics o Power BI entre otras, siendo esta última es la elegida para este proyecto.

Esta herramienta elegida de Power BI está compuesta por dos versiones, una publicada en internet, y una versión de escritorio. Ambas están en la nube y nos permite depositar y reportar online dashboards con otros participantes. Ambas son accesible desde cualquier buscador en la red y la versión móvil también sobre sistemas Ios, Android y Windows. Existen diferentes perfiles con los que dar acceso a los participantes, según sus funciones o roles en la participación y tratamiento de los datos para elaborar el modelo, o simplemente dar acceso a la visualización para la toma de decisiones.

Es de última generación, capaz de analizar grandes cantidades de datos y convertirlos en informes de gran riqueza visual útiles para detectar patrones o características que, de otro modo, sería muy costoso de conseguir, con la ventaja añadida de no necesitar especialistas en informática para su uso y mantenimiento. Cada vez va a ser más necesario el manejo de datos cualquiera que sea nuestra posición en la empresa. Si ésta crece, los datos también y detrás de estos hay una historia que hay que entender, lo que nos permitirá tomar decisiones adecuadas en el momento adecuado. La visualización de datos con matrices de correlación, gráficos de área, diagramas, histogramas, nos permitirá establecer relaciones en principio ocultas con el objetivo de tratar de establecer tendencias de los datos analizados. Con esta visualización y exploración de datos lograremos establecer relaciones entre datos que en principio no creíamos relacionados.

Power BI está compuesta por diferentes herramientas que la integran: Power Pivot, Power Query, Power Map y Power View. Se utiliza el lenguaje Data Analysis Expression (DAX), comúnmente conocido como lenguaje de fórmulas.

En este proyecto se va a desarrollar el Dashboard en Power BI enfocado desde el punto de vista del área comercial. Nos va a permitir conocer que, quien, como, por qué, y cuando consumen nuestros productos los clientes, y de dónde son nuestros clientes, es decir, cómo se distribuyen nuestras ventas. Para ello se elaborará un modelo de datos: Conjunto de tablas, relacionadas con medidas, jerarquías y KPIs para analizar, todo ello orientado a BI.

Estas son algunas de las ventajas de implementar la visualización de nuestros datos en Power BI, y comunes a todas las empresas:

- 1- Bajo coste.
- 2- Garantía de trabajar con datos fiables y universales.
- 3- Baja o nula probabilidad de manipulación errónea de los datos originales.
- 4- Uso muy intuitivo gracias a la interacción gráfica desde paneles.
- 5- Gran versatilidad y dinamismo en las consultas con el simple uso del ratón.
- 6- Filtrado de datos de forma gráfica sin necesidad de introducir complejas fórmulas o filtros.

En la Ilustración 8 se muestra el Ciclo de Business Intelligence de Microsoft seguido en el proyecto.

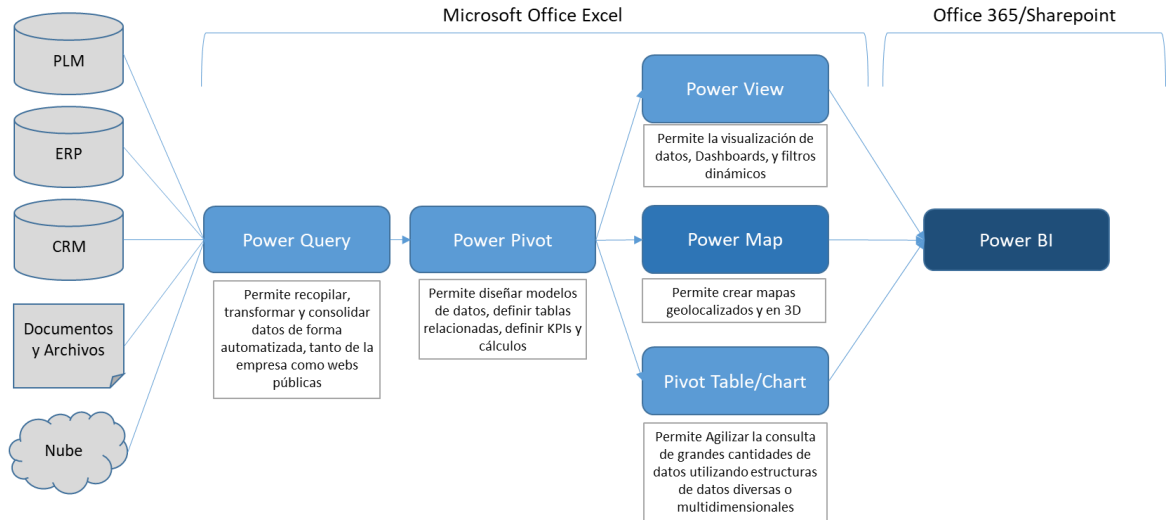


Ilustración 8: Ciclo Business Intelligence de Microsoft. Fuente: SAS Website

Además, se ha creado un mapa básico gracias a Power Map para asociar información cuantitativa y de categorías con ubicaciones espaciales. Así, vemos como se muestra en la Ilustración 9, distribuidas nuestras ventas por código postal del cliente.

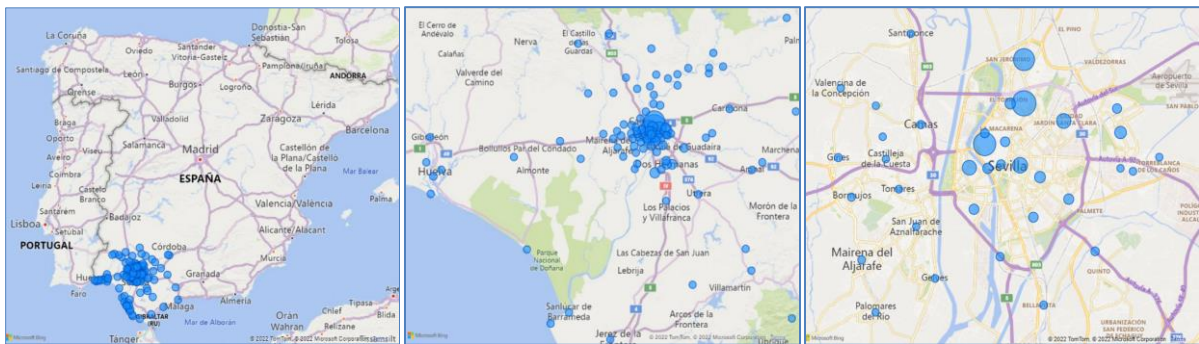


Ilustración 9: Distribución Clientes con Power BI y Power Maps. Elaboración Propia

Power Map nos propone graficar en el mapamundi geográfico, sobre el que incorporamos los datos como se hace en una tabla dinámica, eligiendo los datos a mostrar. La potencia gráfica es enorme de forma que permite por ejemplo ver la distribución de las ventas por lugar y por producto, incluso incorporando la fecha en la barra de reproducción permite ver la evolución de las ventas de forma gráfica de las ventas de cada producto en cada zona, de forma que dando un paseo vemos como un gráfico de barras en distintas zonas sobre un plano evoluciona año tras año, al estilo ya que todo entra por los ojos y de forma muy dinámica, podemos ver la mejor zona de ventas dentro de una provincia, y la peor en un solo minuto.

Estas son algunas de las ventajas obtenidas con la creación del Dashoboar operativo comercial y concretamente con el mapa de clientes creado con Power Map:

- 1- Es una herramienta clave en el seguimiento del plan operativo de la empresa.
- 2- Se han obtenido reportes dinámicos, siendo totalmente manipulable por otros usuarios sin riesgo de pérdida de datos ni de comprensión.

- 3- Ya no hablamos de «datos», hablamos de «información» susceptible de usar en el análisis y toma de decisiones. Conseguimos desglosar los datos sin procesar y demostrar todo mediante tablas, gráficos, videos y más
- 4- Facilitar una visión general del negocio compartida por todos los integrantes de la organización, desde los especialistas en marketing hasta los científicos de datos.
- 5- Ayuda a la toma de decisiones estratégicas a otras áreas de la empresa, no solo la comercial: Marketing, comunicación, logística y distribución, etc.

### 4.3 Entrenamiento y aplicación del modelo ML basado en FCA

Para el siguiente paso de aplicación práctica, necesitamos generar una matriz de doble entrada de clientes y productos adquiridos, en la que los valores posibles son 1 si el cliente adquirió al menos 1 unidad del producto y 0 en caso contrario. Acotaremos el periodo temporal al año 2020 por ser el más reciente disponible.

Creamos nuestro contexto con los datos de entrada y el método FCA, representado en forma de matriz, de modo que cada fila representa un objeto del dominio de interés, y cada columna representa uno de los atributos definidos. Los elementos de la matriz de entrada solo pueden asumir Valores booleanos, es decir, cualquier objeto tiene o no tiene un atributo particular. Si un objeto tiene un atributo particular se coloca una marca (por ejemplo, el símbolo "X") en la intersección de la fila de ese objeto y la columna de ese atributo. Esto se muestra en la Ilustración 10 a continuación:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1		0	10	10042	10046	10048	10052	10054	10055	10062	10066	10067	10069	10070	10073	100
2	155	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	217	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0
4	1509	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	1835	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	2195	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	2303	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
8	2698	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0
9	3240	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
10	3559	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
11	3979	0	0	0	3	0	0	0	0	0	0	0	0	2	0	0
12	4640	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13	4769	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	5555	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	5587	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16	5858	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17	6672	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
18	7279	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19	8154	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
20	8494	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
21	8713	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
22	9040	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
23	9211	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0
24	9220	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Ilustración 10: Matriz binaria, como input al FCA. Elaboración Propia

De lo contrario, si un objeto no tiene un determinado atributo, la intersección de la fila de ese objeto y la columna de ese atributo quedan vacíos. Esta matriz de entrada se define como un contexto formal sobre el cual se realizará el análisis.

La matriz obtenida de la tabla procesada contiene casos donde el nº de unidades vendidas para un producto y cliente dados puede ser mayor de uno, estos casos se normalizan posteriormente al valor "1" ya que se requiere que la matriz tenga valores binarios para el paso posterior de generación de reglas.



Una vez disponemos de la información estructurada de esta forma, queremos extraer las reglas (implicaciones) que se pueden derivar según el modelo FCA. Para ello usaremos la herramienta *conexp* en su versión 1.3.

Conexp (Concept Explorer) es una herramienta utilizada en la investigación y aplicación del *Formal Concept Analysis*. Permite la creación y exploración de tablas simples de objetos y atributos (el tipo de estructura de información a la que hemos llegado en el paso anterior), llamadas “contextos” en FCA, y la extracción de implicaciones y reglas de asociación que se derivan de los mismos.

En nuestro caso, abrimos la estructura binaria que hemos creado (formato csv es válido) y la visualización nos muestra de una forma similar a tabla de doble entrada las relaciones entre clientes y productos, mostrada en la Ilustración 11 a continuación:

The screenshot shows the Concept Explorer interface. On the left, there is a 'Document' pane with a tree view containing 'compras2020...' and 'Context'. Below it is a 'Parameter' table:

Parameter	Value
Show arrow relati...	don't show
Compressed	<input type="checkbox"/>
Object count	1135
Attribute count	75

The main area is a matrix with columns labeled A through N and rows labeled with object IDs. The matrix contains 'X' marks indicating relationships between objects and attributes.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
155														
217		X												
1509														
1835														
2195														
2303														
2698											X	X		
3240											X	X		
3559											X			
3979					X									X
4840														
4769														
5555														
5587														
5858														
5672													X	
7279														
8154														
8494														
8713														
9040														
9211		X										X		
9220														
9253														
9274														
9444		X									X		X	
9669														
10145														
10329		X									X			
10600														
10897														
11014		X									X			
11040														
11824														
12314													X	
12334														
12477		X												

Ilustración 11: Matriz de Entrada en Concep Explorer del FCA. Elaboración Propia

Esto te permite visualizar las reglas de asociación existente entre los datos de entrada, y se muestran a continuación en la Ilustración 12:

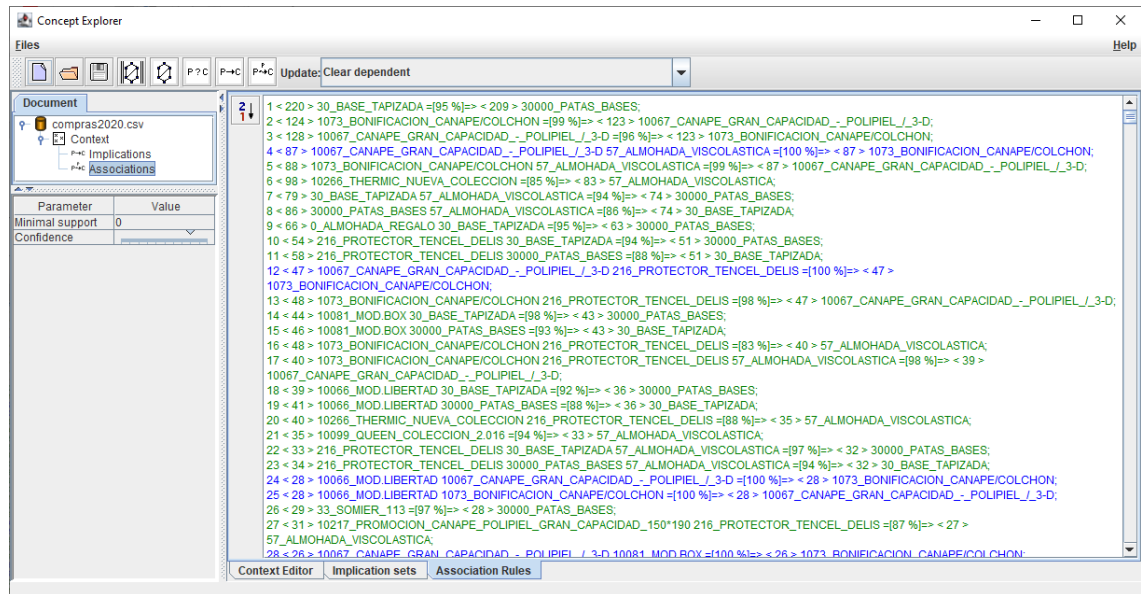


Ilustración 12: Reglas de Asociación en Concep Explorer. Elaboración Propia

A través de los botones de la barra superior, el programa nos permite visualizar el retículo de conceptos (lattice) resultante de esta tabla o contexto, el conjunto de implicaciones (*set of implications*) y el listado de reglas de asociación (*association rules*).

En particular, son estas reglas de asociación en las que nos basaremos para realizar recomendaciones. Estas siguen una lógica de antecedentes-consecuentes.

Dado que en el siguiente paso usaremos un código específico en Java para la extracción de las reglas de asociación, nos limitaremos a guardar el modelo como un fichero de formato nativo de *conexp* (.cxt).

Como decíamos, ahora generaremos un fichero de texto que nos recopile las reglas de asociación del contexto de datos, haciendo una llamada mediante un script de Java y extrayendo únicamente esas reglas del fichero que previamente hemos guardado con *conexp*.

```
java -jar FCA_Solo_Reglas.jar compras2020.cxt >reglas.txt
```

Script en Java para extracción de las reglas
Fichero donde se encuentran las reglas
Fichero output

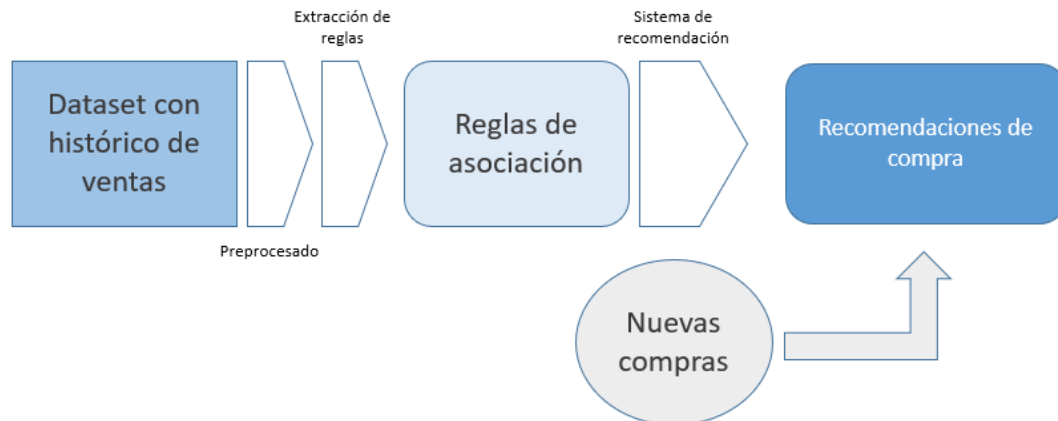
El resultado es un txt con un contenido similar al texto de la captura anterior, que nos será necesario para tratar estas dependencias en nuestro sistema de recomendación.

En el siguiente paso crearemos nuestro sistema de recomendación. Usaremos el lenguaje python para escribir un código cuyo input serán las reglas de asociación y cuyo output será un listado de productos sugeridos para un cliente dado.

Lo que determinará que los productos recomendados sean unos u otros será combinación de un input “genérico” (las reglas) y de un input “particular” (a la hora de evaluar el cliente, el modelo revisa los productos que el cliente ya ha adquirido). En esencia, el sistema busca qué productos

han comprado clientes con compras similares al cliente evaluado y, en su defecto, se basa en los productos con mayor frecuencia de venta en la población total de clientes.

En la Gráfica 7 se detalla la metodología utilizada para la generación del modelo de recomendación de compras, partiendo del dataset histórico de ventas de los clientes.



Gráfica 7: Modelo de generación de recomendaciones. Elaboración Propia

Aquí nos encontramos con una de las limitaciones de los sistemas de recomendación de tipo colaborativo: si el cliente evaluado aún no tiene histórico de compras, el resultado puede no ser satisfactorio al recurrir a los gustos del total de la población (a falta de disponer de atributos de cliente que pudiera permitir recomendar lo que han comprado perfiles de cliente similar, sin necesidad de tener adquisiciones en común).

Sin embargo, como ventajas de este tipo de sistemas se encuentran la sencillez y transparencia en su funcionamiento; de hecho, no necesitaremos ninguna librería específica, sino que usaremos operaciones matemáticas básicas existentes por defecto en el lenguaje.

En un primer bloque del script a crear, definiremos:

- Las funciones para cargar de forma adecuada como cadenas de texto tanto el input general (las reglas) como el input genérico (los históricos de compra de los clientes a los que realizar recomendación).
- La función principal que realizará las recomendaciones.

En el segundo bloque, escribimos el programa principal, en el que se definen como parámetros del script los ficheros txt donde se encuentran las reglas y los históricos de compra de los clientes a evaluar, respectivamente.

La función `cargar_reglas()` recorre línea a línea del txt de las reglas de asociación y las va agregando en un elemento lista, separandolas previamente en los diferentes elementos que las componen, separados por “;”, apoyándose en la función `analizar()`.

La función `analizar()` estructura una regla de asociación como una tupla donde cada elemento es un valor número o una lista de caracteres, y tiene un significado específico dentro de la regla (soporte, confianza, antecedentes, consecuentes).

Por ejemplo, la primera línea del fichero de reglas de asociación es:

```
1;100;10042_CANGURO_113,10220_MOD_ANIVERSARIO_RUPE_1.970;30000_PATAS_BASES,10
098_BROOKLYN_(NéCLEO_AMERICANO),33_SOMIER_113,57_ALMOHADA_VISCOLASTICA
```

La función analizar() la estructura como:

*soporte*

*confianza*

*antecedentes*

*consecuentes*

```
(1,
100,
['10042_CANGURO_113', '10220_MOD_ANIVERSARIO_RUPE_1.970'],
['30000_PATAS_BASES',
'10098_BROOKLYN_(NéCLEO_AMERICANO)',
'33_SOMIER_113',
'57_ALMOHADA_VISCOLASTICA'])
```

Y la función cargar\_reglas() va agregando sucesivamente cada una de las reglas una vez leídas con esa estructura en el elemento lista anteriormente mencionado:

Regla 1 { [(1,  
100,  
['10042\_CANGURO\_113', '10220\_MOD\_ANIVERSARIO\_RUPE\_1.970'],  
['30000\_PATAS\_BASES',  
'10098\_BROOKLYN\_(NéCLEO\_AMERICANO)',  
'33\_SOMIER\_113',  
'57\_ALMOHADA\_VISCOLASTICA']),

Regla 2 { [(1,  
100,  
['10042\_CANGURO\_113', '36\_BONIFICACION\_FIDELIZACION\_CLIENTE'],  
['10081\_MOD.BOX',  
'1073\_BONIFICACION\_CANAPE/COLCHON',  
'57\_ALMOHADA\_VISCOLASTICA',  
'10067\_CANAPE\_GRAN\_CAPACIDAD\_-\_POLIPIEL/\_3-D',  
'10080\_MOD.VISCO-SOFT'])]

Recorre línea a línea el fichero con los históricos de compra de los diferentes clientes y genera con ellos una lista de listas:

Ejemplo:

```
10081_MOD.BOX
10067_CANAPE_GRAN_CAPACIDAD_-_POLIPIEL/_3-D,10099_QUEEN_COLECCION_2.016
xx
```

Con esto, se obtiene los siguientes resultados o salida de la función con lo recomendado:

```
[[['10081_MOD.BOX'], ['10067_CANAPE_GRAN_CAPACIDAD_-_POLIPIEL/_3-D', '10099_QUEEN_CO
LECCION_2.016'], [''], ['xx']]]
```

En la tercera línea se incluye un término genérico “xx” para representar el funcionamiento del modelo cuando el histórico de compra no coincide con ningún otro cliente de la información pasada.

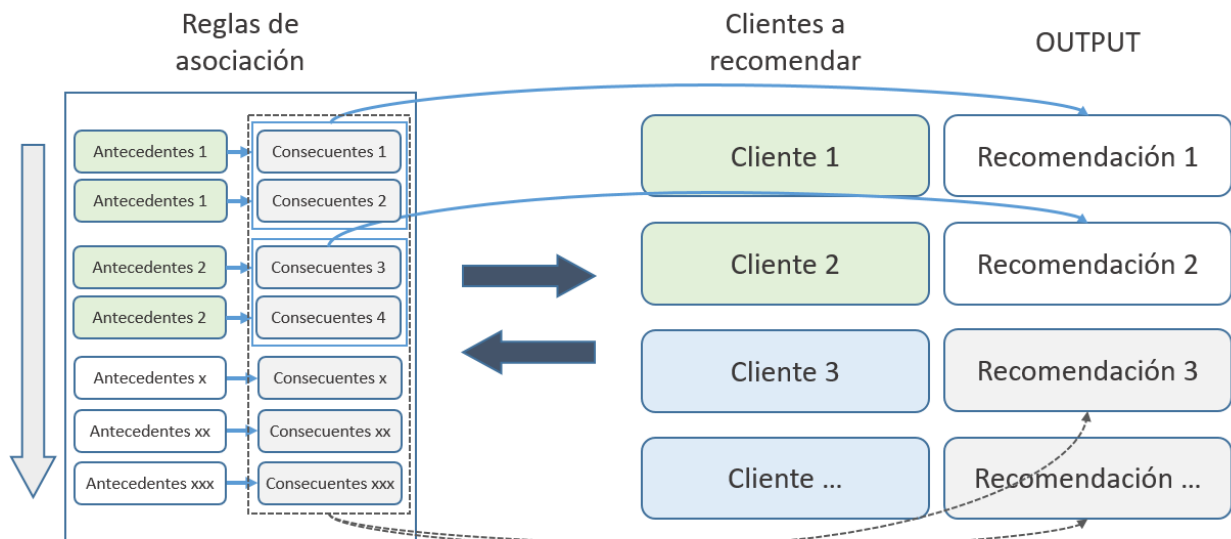
La función de recomendación realiza efectivamente la evaluación de qué productos se recomiendan para cada perfil teniendo en cuenta su histórico de compra.

En detalle, a esta función se le proporcionan como inputs: 1. las reglas inferidas del histórico de ventas y 2. las compras previas de los clientes a los que recomendar. Con ello se crea una lista vacía donde se guardará la recomendación para cada cliente y se lanza un bucle que recorre cada regla: se toma su antecedente (las compras previas a una siguiente decisión de compra).

Se comprueba si coincide con las compras previas de alguno de los clientes a recomendar, en cuyo caso tomaremos su consecuente (compras posteriores). En caso de que no exista esa coincidencia, imputará como resultado a recomendar todos los consecuentes contenidos en el conjunto de las reglas.

En la parte final del código se ordenan los resultados de mayor a menor en función del índice de confianza, y se devuelve los primeros resultados contenidos en la lista final (en este caso, el código está limitado a 5 recomendaciones).

De esa forma, la recomendación para un cliente cuyas compras previas aparecen en las reglas serán los productos comprados con mayor frecuencia en el conjunto de consecuentes seleccionados, y de lo contrario se tomarán los productos del conjunto total de consecuentes. Todo esto se relaciona en la Gráfica 8, a continuación:



Gráfica 8: Esquema de relación de los antecedentes existentes con el perfil de histórico de compras o, en su defecto, con el genérico de los consecuentes. Elaboración Propia.

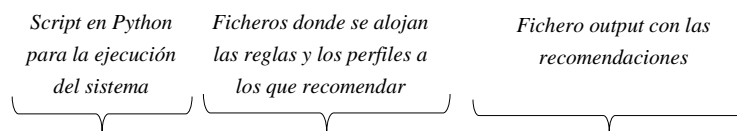
Una vez hemos desarrollado las diferentes funciones necesarias para que el sistema de recomendación funcione (desde la carga de las reglas, los perfiles a los que recomendar, y la lógica del modelo), vamos a ensamblar las diferentes partes para que pueda ser ejecutado como un todo en forma de script y alimentado por ficheros input externos donde estén alojadas las reglas y los perfiles.

Creamos el programa que se compondrá de 3 parámetros: el primero será el fichero “script” donde se aloje el propio código, el segundo será el fichero con las reglas y el tercero el fichero con los perfiles de compra a los que realizar la recomendación:

Este código lo guardaremos como un fichero en formato .py para que sea formato script.

A su vez, creamos un fichero de texto que realice la llamada a ese script de Python y a los 2 parámetros que indicábamos anteriormente, marcándole como salida un fichero .txt (“ficherosugerencias.txt”) donde se almacenarán las recomendaciones realizadas por el modelo:

```
python sugerencias.py reglas.txt compras.txt >ficherosugerencias.txt
```



Y la siguiente salida del programa, con los resultados obtenidos:

```
[ '10081_MOD.BOX' ] ==> [ ('57_ALMOHADA_VISCOLASTICA', 45), ('0_ALMOHADA_REGALO', 45), ('57_ALMOHADA_VISCOLASTICA', 37), ('0_ALMOHADA_REGALO', 31), ('10076_CANAPE_RUPNATURE_ESTILO', 0) ]
[ '10067_CANAPE_GRAN_CAPACIDAD_-POLIPIEL_/_3-D', '10099_QUEEN_COLECCION_2.016' ] ==> [ ('1073_BONIFICACION_CANAPE/COLCHON', 100), ('57_ALMOHADA_VISCOLASTICA', 100), ('1073_BONIFICACION_CANAPE/COLCHON', 96), ('57_ALMOHADA_VISCOLASTICA', 94), ('216_PROTECTOR_TENCEL_DELIS', 45) ]
[ 'xx' ] ==> [ ('57_ALMOHADA_VISCOLASTICA', 37), ('0_ALMOHADA_REGALO', 31), ('10076_CANAPE_RUPNATURE_ESTILO', 0), ('96_RELLENO_MOD._ALFA-2002', 0), ('4_CANAPE_MADERA_-PINTADO_CARTA_DE_COLORES_(EXTINGUIDO)', 0) ]
```

La salida del programa es, para cada perfil de compra dado (con uno o varios productos comprados), una lista de 5 productos distintos ordenados de mayor a menor índice de recomendación. Recordemos que, en caso de no haber 5 productos coincidentes en las reglas de asociación que coincidan con el perfil al que recomendamos, se toman los productos de mayor frecuencia en el resto de la población.

El último de los tres perfiles de compra se corresponde precisamente con un caso genérico que, al no encontrarse en ninguna de las reglas, devuelve un listado que refleja completamente recomendaciones extraídas del total de la población.

En conclusión, este método de elaboración de un sistema de recomendación basado en FCA permite extraer las reglas lógicas mediante un sistema “formal” (como propiamente indica el nombre “Formal Concept Analysis”, y aplicarlas según una lógica conocida y transparente, a diferencia del funcionamiento de otros sistemas como puedan ser las redes neuronales, cuyo

funcionamiento es opaco, tipo “caja negra”. Además, la sencillez de su construcción permite que se pueda aplicar y escalar a multitud de casos de uso y conseguir resultados en un menor tiempo.

El soporte y la confianza son dos métricas utilizadas para medir la importancia y la relevancia de una regla de asociación.

El soporte de una regla de asociación se refiere al porcentaje de transacciones en las que aparece la regla. Por ejemplo, si la regla "si un cliente compra un producto A, es muy probable que también compre un producto B" aparece en el 10% de las transacciones, su soporte sería del 10%. El soporte se utiliza para medir la frecuencia con la que aparece la regla en los datos.

La confianza de una regla de asociación se refiere al porcentaje de veces que se cumple la regla. Por ejemplo, si la regla "si un cliente compra un producto A, es muy probable que también compre un producto B" se cumple en el 80% de las veces que aparece en los datos, su confianza sería del 80%. La confianza se utiliza para medir la fuerza de la relación entre los conceptos formales de la regla.

El soporte y la confianza son métricas muy útiles para evaluar la relevancia y la importancia de una regla de asociación en el contexto de un sistema de recomendación. Reglas con un alto soporte y una alta confianza son más fiables y tienen más probabilidades de ser útiles para el sistema de recomendación.

#### 4.4 Validación de datos

Es importante medir la calidad de nuestro sistema de recomendación realizado, antes de integrarlo en la empresa. En la Tabla 4, se representan las diferentes combinaciones de recomendaciones que se pueden generar en una recomendación. Una recomendación es positiva si la calificación recomendada coincide con los gustos de los cliente y es realmente una venta potencial de compra. Para medir la calidad del sistema de recomendación, se utilizan métricas de precisión y tasa de acierto, ampliamente utilizados por RS para evaluar su recomendación. (Das et al., 2014)

	<b>Compras Potenciales</b>	<b>Compras No Potenciales</b>
<b>Recomendados</b>	Verdadero Positivo	Falsos Positivos
<b>No Recomendados</b>	Falsos Negativos	Verdadero Negativo

Tabla 4 : Calidad de resultados en las recomendaciones. Elaboración Propia

La precisión mide el grado de exactitud de las recomendaciones producidas por el algoritmo. En nuestro sistema, la precisión mide qué fracción de la recomendada los artículos son potenciales de compra por el cliente.

$$Tasa\ de\ precisión = \frac{Verdadero\ Positivo}{Verdadero\ Positivo + Falsos\ Positivos}$$

La tasa de acierto mide qué fracción de los artículos que gustan a los clientes ha sido recomendado por el algoritmo.

$$Tasa\ de\ acierto = \frac{Verdadero\ Positivo}{Verdadero\ Positivo + Falsos\ Negativo}$$

Posteriormente, podremos añadir más métricas de medición de la calidad introduciendo al sistema de recomendación compras posteriores en las que ya se realizó la recomendación al cliente.

Para ello, se hace una nueva extracción de las reglas a partir de una muestra aleatoria del 80% de los registros de clientes que compraron en 2020, reservando el 20% restante como conjunto de test. Estos registros no los verá el modelo para la extracción de reglas.

Es habitual también realizar una validación cruzada, que consiste en entrenar al modelo varias veces usando el conjunto de entrenamiento y excluyendo en cada ocasión un set distinto de dicho conjunto, siendo el acierto la media de cada una de las iteraciones. El inconveniente de este sistema es el mayor tiempo y esfuerzo computacional dedicado al entrenamiento, por ello realizaremos solo un entrenamiento con el 80% de los registros.

Una vez entrenado el modelo y tenemos las reglas de asociación, disponemos del 20% de los perfiles de compra (227 casos) para medir la calidad de las recomendaciones. En este punto debemos tener en cuenta la lógica de negocio y las peculiaridades de los datos utilizados para realizar la validación de la forma más adecuada.

Cada uno de los perfiles de compra (clientes) han podido comprar uno o más productos en el periodo temporal de 2020. Para que el RS pueda generar recomendaciones debemos pasarle alguna de las compras de cada uno de los perfiles como input. Por ello determinamos que el input serán 2 productos comprados, y con ello compararemos las recomendaciones con el resto de artículos comprados por el cliente.

Como consecuencia, de los 227 clientes en la partición de test, podemos quedarnos con 120 que son los que tienen al menos 3 compras (2 serán input o variable independiente y el resto, al menos 1, será la variable dependiente que compararemos con las predicciones de la RS).

Finalmente, definimos las métricas de medición de la calidad que utilizaremos. Respecto a la tasa de precisión, esta va a estar marcada por el número de recomendaciones que hayamos pedido al RS para cada perfil de compra. A mayor número de ellas, es probable que la tasa de precisión disminuya ya que estaremos aumentando el nº de falsos positivos, es decir, estaremos considerando demasiadas recomendaciones potenciales para un nº de compras reales finitas. En nuestro RS hemos limitado a 5 el nº de recomendaciones por cliente

$$Tasa\ de\ precisión = \frac{Recomendaciones\ que\ fueron\ compra\ real}{Rec.\ compra\ real + \underbrace{Rec.\ no\ compra\ real}_{Tipo\ de\ recomendaciones\ a\ minimizar}}$$

En cuanto a la tasa de acierto, dada la naturaleza de la información se nos presentan 2 posibles variantes: tasa de acierto a nivel cliente (se considera acierto si alguna de las compras reales del cliente estaba entre las recomendaciones) y tasa de acierto a nivel producto (se considera un acierto por cada uno de los productos comprados realmente y que hubieran sido recomendados por el RS).



$$\begin{aligned} \text{Tasa de acierto (cliente)} &= \\ &= \frac{\text{Clientes } \geq 1 \text{ compra fue recomendada}}{\text{Clientes } \geq 1 \text{ compra fue recomendada} + \text{Clientes 0 compras fueron recomendadas}} \end{aligned}$$

$$\text{Tasa de acierto (productos)} = \frac{\text{Compras recomendadas}}{\text{Compras recomendadas} + \text{Compras no recomendadas}}$$

Con ello, el resumen de resultados obtenido se muestra en la Tabla 5 a continuación:

	Unidad: Clientes	Unidad: Productos			
	Compras Potenciales	Compras Potenciales	Compras No Potenciales	TOTAL	TASA DE PRECISIÓN
<b>Recomendados</b>	90	138	234	<b>372</b>	<b>37,1%</b>
<b>No Recomendados</b>	30	108			
<b>TOTAL</b>	<b>120</b>	<b>246</b>			
<b>TASA DE ACIERTO</b>	<b>75,0%</b>	<b>56,1%</b>			

Tabla 5: Resumen resultados del sistema de Recomendación. Elaboración Propia

De lo que se desprende que el 75% de los clientes adquirieron al menos 1 producto de los que el RS le habría recomendado, el 56% de los productos comprados habrían sido recomendados por el RS, y que el 37% de las recomendaciones coincidieron con el comportamiento real del cliente.

Adicionalmente vemos que el nº de recomendaciones es ligeramente mayor al nº de compras realizadas (siendo esto lo deseable, de lo contrario supondría que estamos haciendo poco esfuerzo de promoción en relación a la cifra de ventas actual). Incluso sería aconsejable incrementar más la proporción recomendaciones / ventas, parametrizando un mayor nº de recomendaciones a cada cliente (recordemos que se han generado 5 para cada uno, pero el número real es menor ya que se generan duplicados en algunos casos).

## 4.5 Visualización de los resultados

Lo ideal en este apartado sería disponer de una interfaz gráfica que nos mostrara los resultados integrado en su software, pero esto queda fuera del alcance del proyecto.

Para este proyecto los veremos a través del archivo de texto que genera nuestro sistema de recomendaciones. Gracias a los outputs de nuestro Sistema de Recomendación, disponemos entre otros de un reporte de estadísticas sobre el grado de adopción de las recomendaciones por parte de los clientes, métricas de: Tasa de acierto, conversión de las recomendaciones, tasa de visualización de las recomendaciones, datos de clics a los mails que se comunican, con que medios se convierte más, segmentación de los clientes con mayor conversión, etc.

La implementación de este algoritmo iría integrado con la base de datos de la empresa y con el software de facturación, que aplicaría el código implementado en este proyecto directamente sobre los inputs del cliente que realiza la compra, la procesaría y leería del servidor las ventas del resto de histórico de ventas almacenadas para devolver al cliente los productos o servicios que les recomendaría directamente por pantalla (o por otro medio como mail, SMS, etc. según lo programemos).



## 5 Resultados y conclusión

El Sistema de Recomendación creado, pronostica y clasifica correctamente las características de los clientes para crear un modelo estratégico crucial para la organización, ayudando además en la toma de decisiones estratégicas para aumentar sus potenciales ventas y beneficios.

Por tanto, concluimos que todos los objetivos propuestos en este proyecto han sido alcanzados, y a continuación describimos su cumplimiento.

Los resultados específicos presentados e ilustrados muestran todas las dimensiones del análisis de datos que no solo mejoran significativamente las ventas de las empresas, sino que también demuestran cómo la competencia de análisis de datos se relaciona con el desempeño en la toma de decisiones.

Además, este trabajo logra demostrar la aplicabilidad y superioridad del método/modelo creado sobre los anteriores existentes en la empresa (basados en la intuición y experiencias propias).

La información proporcionada es altamente modular y permite la futura inclusión de nuevas funciones para proporcionar análisis de datos más grandes y mejorados. Para cada cliente que introducimos en el modelo, y según a sus patrones de compras, características y productos adquiridos, el modelo nos devuelve satisfactoriamente el número de productos a comprar que le indicamos que queremos que nos recomiende, todo ello basado en la inteligencia artificial.

Además, otro aporte fundamental de este proyecto es la baja inversión necesaria para la empresa para su implementación. Es una herramienta que posibilita la mejora de su estrategia comercial a partir de los datos e información de las que dispone actualmente, sin necesidad de adquirir costosos softwares o tecnologías.

Este proyecto puede ser interesante para una audiencia de universidades que estudian o investigan temas similares dentro de los Sistemas de Recomendación, además del sector de la inteligencia Artificial (IA).



## 6 Futuros proyectos y líneas de investigación

Finalmente, diferentes desafíos podrían ser relevantes para continuar mejorando el trabajo, por lo que ahora damos algunas recomendaciones para futuras investigaciones.

Desde la perspectiva no solo de los autores sino también de la empresa y la universidad, se elabora un modelo interesantes y de gran valor añadido tanto para la práctica, como para la labor académica. Esto colabora con el propósito de abrir futuras líneas de investigación, o aportar a las ya existentes. Así como la aplicación práctica de las múltiples teorías y fundamentos documentados a lo largo de la literatura del aprendizaje automático.

Por ello, las contribuciones del estudio ahora se pueden analizar junto a sugerencias para futuras investigaciones dentro del tema.

Este modelo también se puede validar en el futuro con otras técnicas de IA, no FCA, para garantizar resultados precisos que se pueden implementar en un sistema de la vida real.

Debido a las limitaciones del software FCA, el enfoque creado también se puede reconstruir utilizando otro software de IA que, aunque es menos fácil de usar, tiene muchas más opciones de configuración y, por lo general, es más exigente desde el punto de vista computacional para la optimización.

Como próximo paso, se inicia una línea trabajo para crear una herramienta de telemarketing de análisis offline para mejorar el servicio de postventa de nuestros clientes, utilizando nuestro modelo de SR construido.



## Bibliografía

1. Ahn, J. S., & Sohn, S. Y. (2009). Customer pattern search for after-sales service in manufacturing. *Expert Systems with Applications*, 36(3), 5371-5375.
2. Armstrong, W. W. (1974, August). Dependency Structures of Data Base Relationships. In *IFIP congress* (Vol. 74, pp. 580-583).
3. Aranda-Corral, G. A., Borrego Díaz, J., & Galán Páez, J. (2011, June). Confidence-based reasoning with local temporal formal contexts. In *International Work-Conference on Artificial Neural Networks* (pp. 461-468). Springer, Berlin, Heidelberg.
4. Belohlavek, R. (2008). Introduction to formal concept analysis. *Palacky University, Department of Computer Science, Olomouc*, 47.
5. Bertolini, M., Mezzogori, D., Neroni, M., & Zammori, F. (2021). Aprendizaje automático for industrial applications: A comprehensive literature review. *Expert Systems with Applications*, 175, 114820.
6. Boucher-Ryan, P. D., & Bridge, D. (2005, December). Collaborative recommending using formal concept analysis. In *International Conference on Innovative Techniques and Applications of Artificial Intelligence* (pp. 205-218). Springer, London.
7. Carlson, J. (2014). The use of life cycle models in developing and supporting data services. *Research data management: Practical strategies for information professionals*, 63-86.
8. Chemmalar Selvi, G., Lakshmi Priya, G. G., & Joseph, R. B. (2019). A FCA-based concept clustering recommender system. In *Context-Aware Systems and Applications, and Nature of Computation and Communication* (pp. 178-187). Springer, Cham.
9. Corral, G. A. A. (2007). Tratamiento de la semántica emergente mediante sistemas de agentes basados en el conocimiento. In *I Simposio de Doctorado en Web Semántica: Salamanca. 13 de noviembre de 2007* (pp. 10-19). Universidad de Salamanca.
10. Das, J., Mukherjee, P., Majumder, S., & Gupta, P. (2014, November). Clustering-based recommender system using principles of voting theory. In *2014 International conference on contemporary computing and informatics (IC3I)* (pp. 230-235). IEEE. [B]
11. Ebrahimi, B., Dellnitz, A., Kleine, A., & Tavana, M. (2021). A novel method for solving data envelopment analysis problems with weak ordinal data using robust measures. *Expert Systems with Applications*, 164, 113835.
12. Ganter, B., & Wille, R. (2012). *Formal concept analysis: mathematical foundations*. Springer Science & Business Media.
13. Ganter, B., Stumme, G., & Wille, R. (Eds.). (2005). *Formal concept analysis: foundations and applications* (Vol. 3626). Springer.
14. Gomez-Urbe, C. A., & Hunt, N. (2015). The netflix recommender system: Algorithms, business value, and innovation. *ACM Transactions on Management Information Systems (TMIS)*, 6(4).
15. Jorro-Aragoneses, J. L., Caro-Martínez, M., Díaz-Agudo, B., & Recio-García, J. A. (2020, June). A user-centric evaluation to generate case-based explanations using formal concept analysis. In *International Conference on Case-Based Reasoning* (pp. 195-210). Springer, Cham
16. Kuznetsov, S. O., & Schmidt, S. (Eds.). (2007). *Formal Concept Analysis: 5th International Conference, ICFA 2007, Clermont-Ferrand, France, February 12-16, 2007, Proceedings* (Vol. 4390). Springer.
17. Kuznetsov, S. O., & Schmidt, S. (Eds.). (2007). *Formal Concept Analysis: 5th International Conference, ICFA 2007, Clermont-Ferrand, France, February 12-16, 2007, Proceedings* (Vol. 4390). Springer.
18. Maddodi, S. (2019). NETFLIX bigdata analytics-the emergence of data driven recommendation. *Srivatsa Maddodi, & Krishna Prasad, K. (2019). Netflix Bigdata Analytics-The Emergence of Data Driven Recommendation. International Journal of Case Studies in Business, IT, and Education (IJCSBE)*, 3(2),
19. Melville, P., & Sindhvani, V. (2010). Recommender systems. *Encyclopedia of aprendizaje automático*, 829-838.
20. Ortega, F., Sanchez, J. L., Bobadilla, J., & Gutierrez, A. (2013). Improving collaborative filtering-based recommender systems results using Pareto dominance. *Information Sciences*, 239, 50-61. [D]
21. Pérez-González, C. J., Colebrook, M., Roda-García, J. L., & Rosa-Remedios, C. B. (2019). Developing a data analytics platform to support decision making in emergency and security management. *Expert Systems with Applications*, 120, 167-184.
22. Škopljanač-Maćina, F., & Blašković, B. (2014). Formal concept analysis—overview and applications. *Procedia Engineering*, 69, 1258-1267. [A]
23. Škopljanač-Maćina, F., & Blašković, B. (2014). Formal concept analysis—overview and applications. *Procedia Engineering*, 69, 1258-1267.





## Selección artículos importantes

1. Kataria, S., & Batra, U. (2022). Co-clustering neighborhood—based collaborative filtering framework using formal concept analysis. *International Journal of Information Technology*, 14(4), 1725-1731.
2. Liu, Z. H., Zhao, Q., Zou, L., Xu, W. H., & Min, F. (2022). A heuristic concept construction approach to collaborative recommendation. *International Journal of Approximate Reasoning*, 146, 119-132.
3. Boujemaa, K. S., Berrada, I., Fardousse, K., Naggar, O., & Bourzeix, F. (2021). Toward Road Safety Recommender Systems: Formal Concepts and Technical Basics. *IEEE Transactions on Intelligent Transportation Systems*, 23(6), 5211-5230.
4. Indira, K., Karthiga, S., Nisha Angeline, C. V., & Santhiya, C. (2021). Parallel CLARANS Algorithm for Recommendation System in Multi-cloud Environment. In *Computer Networks and Inventive Communication Technologies* (pp. 461-472). Springer, Singapore.
5. Bellini, E., Iraqi, Y., & Damiani, E. (2020). Blockchain-based distributed trust and reputation management systems: A survey. *IEEE Access*, 8, 21127-21151.
6. Bunnell, L., Osei-Bryson, K. M., & Yoon, V. Y. (2020). RecSys issues ontology: a knowledge classification of issues for recommender systems researchers. *Information Systems Frontiers*, 1377-1418.
7. Chen, Y. H., Lu, E. J. L., & Cheng, Y. W. (2020). Categorization of Multiple Documents Using Fuzzy Overlapping Clustering Based on Formal Concept Analysis. *International Journal of Software Engineering and Knowledge Engineering*, 30(05), 631-647.
8. Cordero, P., Enciso, M., Mora, Á., Ojeda-Aciego, M., & Rossi, C. (2020). A Formal Concept Analysis Approach to Cooperative Conversational Recommendation. *International Journal of Computational Intelligence Systems*, 13(1), 1243-1252.
9. GG, L. P. (2020). Rating Prediction Method for Item-based Collaborative Filtering Recommender Systems Using Formal Concept Analysis. *EAI Endorsed Transactions on Energy Web*, 8(33), e11.
10. GG, L. P. (2020). Three-way formal concept clustering technique for matrix completion in recommender system. *International Journal of Pervasive Computing and Communications*.
11. Jorro-Aragoneses, J. L., Caro-Martínez, M., Díaz-Agudo, B., & Recio-García, J. A. (2020, June). A user-centric evaluation to generate case-based explanations using formal concept analysis. In *International Conference on Case-Based Reasoning* (pp. 195-210). Springer, Cham.
12. Menšík, M., Albert, A., & Patschka, V. (2020). Using FCA for seeking relevant information sources. *Recent Advances in Slavonic Natural Language Processing, 2020*, 47-54.
13. Muangprathub, J., Boonjing, V., & Chamnongthai, K. (2020). Learning recommendation with formal concept analysis for intelligent tutoring system. *Heliyon*, 6(10), e05227.
14. Vilakone, P., & Park, D. S. (2020). The Efficiency of a DoParallel Algorithm and an FCA Network Graph Applied to Recommendation System. *Applied Sciences*, 10(8), 2939.
15. Carbonnel, J., Huchard, M., & Nebut, C. (2019, September). Exploring the variability of interconnected product families with relational concept analysis. In *Proceedings of the 23rd International Systems and Software Product Line Conference-Volume B* (pp. 199-206).
16. Chemmalar Selvi, G., Lakshmi Priya, G. G., & Joseph, R. B. (2019). A FCA-based concept clustering recommender system. In *Context-Aware Systems and Applications, and Nature of Computation and Communication* (pp. 178-187). Springer, Cham.
17. Selvi, G. C., & Priya, G. G. (2019, June). A New Approach for Matrix Completion Using Arrow Relation of FCA in Recommender Systems. In *International Conference on Intelligent Computing and Communication* (pp. 839-846). Springer, Singapore.
18. Mezni, H., & Abdeljaoued, T. (2018). A cloud services recommendation system based on Fuzzy Formal Concept Analysis. *Data & Knowledge Engineering*, 116, 100-123.
19. Naïm, H., Aznag, M., Durand, N., & Quafafou, M. (2016, October). Semantic pattern mining based web service recommendation. In *International Conference on Service-Oriented Computing* (pp. 417-432).
20. Akhmaturov, M., & Ignatov, D. I. (2015, October). Context-Aware Recommender System Based on Boolean Matrix Factorisation. In *CLA* (pp. 99-110).

21. Alqadah, F., Reddy, C. K., Hu, J., & Alqadah, H. F. (2015). Biclustering neighborhood-based collaborative filtering method for top-n recommender systems. *Knowledge and I.S.*, 44(2), 475-491.
22. Bedek, M. A., Kopeinik, S., Prünster, B., & Albert, D. (2015, July). Applying the Formal Concept Analysis to introduce guidance in an inquiry-based learning environment. In *2015 IEEE 15th International Conference on Advanced Learning Technologies* (pp. 285-289). IEEE.
23. Kashnitsky, Y. (2015). Visual analytics in FCA-based clustering. arXiv preprint arXiv:1504.05469.
24. Zhang, W., Du, Y. J., & Song, W. (2015). Recommender system with formal concept analysis.
25. Zhang, W., Du, Y., & Song, W. (2015). Followee recommendation based formal concept analysis in social network. *INTERNATIONAL JOURNAL OF INNOVATIVE COMPUTING INFORMATION AND CONTROL*, 11(4), 1155-1164.
26. Zou, C., Zhang, D., Wan, J., Hassan, M. M., & Lloret, J. (2015). Using concept lattice for personalized recommendation system design. *IEEE systems journal*, 11(1), 305-314.
27. Bau, C.T., Chen, R.C., & Huang, C.Y. (2014). Construction of a clinical decision support system for undergoing surgery based on domain ontology and rules reasoning. *Telemedicine and e-Health*, 460-472.
28. Castellanos, A., García-Serrano, A., & Cigarrán, J. (2014, September). Linked data-based conceptual modelling for recommendation: a FCA-based approach. In *International Conference on Electronic Commerce and Web Technologies* (pp. 71-76). Springer, Cham.
29. Ignatov, D. I., Kaminskaya, A., Konstantinova, N., & Konstantinov, A. (2014, August). Recommender system for crowdsourcing platform witology. In *2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)* (Vol. 1, pp. 327-335). IEEE.
30. Ignatov, D. I., Mikhailova, M., Zakirova, A. Y., & Malioukov, A. (2014, August). Recommendation of ideas and antagonists for crowdsourcing platform witology. In *Russian Summer School in Information Retrieval* (pp. 276-296). Springer, Cham.
31. Senatore, S., & Pasi, G. (2013, March). Lattice navigation for collaborative filtering by means of (fuzzy) formal concept analysis. In *Proceedings of the 28th Annual ACM Symposium on Applied Computing* (pp. 920-926).
32. Bouaud, J., Messai, N., Laouénan, C., Mentré, F., & Séroussi, B. (2012). Eliciting patient patterns of physician non-compliance with breast cancer guidelines using formal concept analysis. In *Quality of Life through Quality of Information* (pp. 477-481). IOS Press.
33. De Maio, C., Fenza, G., Gaeta, M., Loia, V., Orciuoli, F., & Senatore, S. (2012). RSS-based e-learning recommendations exploiting fuzzy FCA knowledge modeling. *Applied Soft Computing*, 12(1), 113-124.
34. Encheva, S. (2012, October). Preferences predictions of learning objects supported by collaborative recommendations. In *International Symposium on Intelligence Computation and Applications* (pp. 145-151). Springer, Berlin, Heidelberg.
35. Li, X., Murata, T. (2012). Exploiting Formal Concept Analysis in Customizing Recommendation for New User and Gray Sheep Problems. *IEEJ Transactions on Electronics Information and Systems*, 782-789.
36. Xu, H. S., & Kuang, G. F. (2012). The development of e-commerce recommendation system by applying formal concept analysis and neural network. *JCIT*, 7(10), 150-157.
37. Zheng, Q., & Wang, Y. (2012). Using formal concept analysis and collaborative filtering to recommendation system in e-commerce. *JCIT*, 7(10), 389-396.
38. Sun, X., & Li, J. (2011, August). Application of recommender system based on domain ontology with modern information technology. In *International Conference on Computer Science, Environment, Ecoinformatics, and Education* (pp. 592-597). Springer, Berlin, Heidelberg.
39. Li, X., & Murata, T. (2010, February). A knowledge-based recommendation model utilizing formal concept analysis and association. In *2010 the 2nd International Conference on Computer and Automation Engineering (ICCAE)* (Vol. 4, pp. 221-226). IEEE.
40. Deng, X., Peng, J., & Huang, H. (2009, October). Research on the fractal company modeling based on competence. In *2009 16th International Conference on Industrial Engineering and Engineering Management* (pp. 2136-2140). IEEE.
41. Cherniahovskaya, L., Nougayeva, K., Shkundina, R., & Muximov, P. (2005, June). Decision support making in business process based on object-cognitive analysis. In *2005 IEEE International Technology Management Conference (ICE)* (pp. 1-4). IEEE.

## Anexos

### Código 1: Preprocesamiento de las tablas de datos en SQL

```
SELECT Tabla_Facturas.*, Tabla_Clientes.*, Tabla_FormasPago.*, Tabla_LíneaFactura.*
FROM Tabla_LíneaFactura INNER JOIN (Tabla_FormasPago INNER JOIN (Tabla_Clientes INNER JOIN
Tabla_Facturas ON Tabla_Clientes.cdcliente = Tabla_Facturas.cdcliente) ON Tabla_FormasPago.cdformapago =
Tabla_Facturas.cdformapago) ON Tabla_LíneaFactura.cdfactura = Tabla_Facturas.cdfactura;
```

### Código 2: funciones de carga de las reglas de asociación en Python

```
def analizar(regla):
    lista = regla.split(";")
    soporte = int(lista[0].strip())
    confianza = int(lista[1].strip())
    antecedentes = lista[2].strip().split(",")
    consecuentes = lista[3].strip().split(",")
    return (soporte,confianza,antecedentes,consecuentes)

def cargar_reglas(camino):
    resultado = [ ]
    f = open(camino)
    for linea in f:
        resultado.append(analizar(linea))
    f.close()
    return resultado
```

### Código 3: funciones de carga de las reglas de asociación: función de carga del histórico de compra de los clientes a los que recomendar en Python

```
def cargar_compras(camino):
    resultado = []
    f = open(camino)
    for linea in f:
        resultado.append([cc.strip() for cc in linea.strip().split(",")])
    f.close()
    return resultado
```

### Código 4: función de recomendación en Python

```
def sugerencias(reglas, compra):
    resultado = []
    for r in reglas:
        cumple = True
        for ant in r[2]:
            if not ant == "":
                cumple = cumple and (ant in compra)
        if cumple:
            for mix in r[3]:
                if not mix in compra:
                    resultado.append((mix,r[1]))
    resultado.sort(key=lambda a: a[1], reverse=True)
    return resultado[:5]
```

**Código 5:** programa principal del sistema de recomendación en Python

```
if __name__ == "__main__" :
    if len(sys.argv) <3:
        print ("Error: Necesita, al menos, 2 parámetros para ejecutar el programa")
        print ("Formato: python sugerencias.py fichero reglas fichero compra ")
    else:
        reglas = cargar_reglas(sys.argv[1])
        compras = cargar_compras(sys.argv[2])
        for c in compras:
            if len(c) == 1 and c[0] == "":
                print ("")
            else:
                if len(c) > 0 :
                    ss = sugerencias(reglas,c)
                    print (c, "\t==>\t",ss)
                else:
                    print ("")
```