



Universidad de Jaén

Facultad de Ciencias Sociales
y Jurídicas

Trabajo Fin de Grado

COMERCIO ELECTRÓNICO Y SISTEMAS DE RECOMENDACIÓN BASADOS EN LA CORRELACIÓN DE PEARSON

Alumno: Victor Jimenez Molina

Julio, 2020

Resumen

En este trabajo de fin de grado estudiaremos la evolución del comercio electrónico. Como la estadística está ayudando a crecer tanto a usuario con sus comprar como a las empresas. Y compararemos la utilización de dos sistemas de recomendación que se utilizan actualmente en páginas de comercio electrónico. Estos métodos se basan en la utilización de coeficientes de correlación de Pearson.

Abstract

In this final year project, we will study the evolution of the e-commerce. How the statistics is helping to grow users with their purchases and the Companies. And we will compare the use of two recommendations systems that are currently used in e-commerce page. These methods are based on the use of Pearson`s correlation coefficients.

Índice

1. Trabajo de fin de Grado.....	3
1.1. Objetivos del trabajo de fin de grado.....	3
1.2. Estructura del trabajo de fin de grado.....	3
2. Comercio electrónico.....	4
2.1. Historia y antecedentes.....	4
2.2. Empresas pioneras.....	6
2.3. El comercio electrónico en la actualidad.....	9
3. Estadística computacional.....	12
3.1. Definición y Concepto.....	12
3.2. Software R.....	13
4. Sistemas de recomendación en comercio electrónico.....	14
4.1. Personalización en páginas webs de comercio.....	14
4.2. Sistemas de recomendación.....	15
4.3. Valoraciones y medidas similitud.....	16
5. Caso de estudio.....	17
5.1. Conjunto de datos de Amazon.....	17
5.2. Filtrado colaborativo basado en usuarios.....	18
5.3. Filtrado colaborativo basado en ítems.....	20
6. Conclusiones.....	22
7. Bibliografía.....	24
8. Anexos: Scripts de R.....	25
8.1. Filtrado colaborativo basado en usuarios.....	25
8.2. Filtrado colaborativo basado en ítems.....	28

Índice de Figuras:

- Figura 1: Conjunto de datos visualizado.....17
- Figura 2: Cuadro de usuarios parecidos a ‘Mike’18
- Figura 3: Cuadro de los productos predichos para ‘Mike’19
- Figura 4: Cuadro productos valorados por el usuario “1234”20
- Figura 5: Comparación productos vistos y no vistos.....20
- Figura 6: Cuadro Top 10 recomendaciones al usuario ‘1234’21
- Figura 7: Script de R del filtrado basado en usuarios.....25
- Figura 8: Scripts de R del filtrado basado en ítems.....28

1. TRABAJO DE FIN DE GRADO

En este apartado explicaremos cuales van a ser los temas a tratar en este trabajo de fin de carrera, cuáles van a ser nuestros objetivos y que estructura vamos a seguir para desarrollarlo.

1.1 Objetivos del trabajo de fin de grado.

Estos son los objetivos que se van a estudiar:

- Explicar que es el comercio electrónico, su historia, su importancia a nivel y que empresas iniciaron esta revolución.
- Explicar que es la estadística computarizada. Que programa estadístico es el más utilizado actualmente y en que se basa.
- Aplicar dos sistemas de recomendación que utilizan la correlación de Pearson compararlos, y ver las ventajas y desventajas de cada uno de ellos.
- Dar una conclusión sobre el tema de estudio.

1.2 Estructura del trabajo de fin de grado.

El trabajo se ha dividido por puntos y apartados. En el primer punto trataremos toda la historia de comercio electrónico. Hablaremos de las dos empresas más importantes de comercio electrónico: Amazon y eBay. También daremos un repaso al comercio electrónico en la actualidad.

Seguidamente hablaremos de la estadística computacional de donde nació y de cómo se utiliza. Hablaremos del software de R dando un breve repaso por su historia y sus creadores.

El punto principal del tfg será hablar de los sistemas de recomendación a usuarios como medida de personalización en las páginas webs. Sobre todo, revisaremos dos de estos sistemas basados en la correlación de Pearson: Basado en las puntuaciones de los usuarios y basados en ítems. Y como parte final expondré las conclusiones respecto al uso de estas técnicas.

Añadiremos un anexo con todos los cálculos realizados en el programa R

2. COMERCIO ELECTRONICO

En este capítulo daremos un repaso por toda la historia del comercio electrónico. Veremos cuales fueron las primeras empresas que empezaron a trabajar de esta forma y como han crecido a lo largo de los años. Finalmente veremos cómo está actualmente esta forma de comercio.

2.1 Historia y antecedentes.

La revolución de la informática y la llamada globalización global han hecho que ahora se tenga nuevos puntos de vista acerca de cualquier aspecto de nuestra vida. Desde poder acercarnos a personas que están a miles de km con un solo clic hasta incluso poder terminar un curso escolar desde nuestra casa, como ha ocurrido con la pandemia de Covid 19 en este año 2020. Y a nadie se le escapa que si no hubiera sido por los avances tecnológicos el mundo hubiera afrontado una crisis casi apocalíptica. Pero a pesar de todo lo que se ha perdido, el teletrabajo y ventanas como el comercio electrónico, se ha podido pasar esta etapa y empezar la llamada “nueva normalidad” desde otro punto de vista.

El comercio electrónico ha sido uno de esos grandes pilares que nos ha permitido desde nuestra casa, poder adquirir cualquier artículo que se nos ocurriese, y a pesar de que algunos servicios de mensajería fueran parados momentáneamente, pero casi todo el servicio se ha mantenido activo. Y permitiendo a las familias poder llevar una normalidad en sus vidas, en la mayoría de lo posible. Pero ¿Qué es el comercio electrónico? ¿Dónde comienza su historia?

Hablamos de comercio electrónico como aquella forma de compra y de venta a través de plataformas de la red.

Pero en cierto modo la historia de este comercio comienza en el año 1920, mucho antes de la invención de internet, cuando se comienza la venta por catálogo. Este sistema era nuevo en su época y no había ningún antecedente hasta el momento. Se compraba sin ver físicamente que se compraba, solo se veía en ese folleto o revista donde se mostraban ilustraciones del producto.

En el año 1960 se inventa un gran sistema de intercambio de datos electrónicos. El sistema EDI (electronic data interchange). Con el objetivo principal de eliminar esas

grandes cantidades de papel y comenzar esa automatización de los procesos que ya llevaban muchos años en otros ámbitos como fabricas etc.... La cantidad de papel que se acumulaba y su difícil gestión y almacenaje conllevaba en muchas ocasiones fallos graves y un gran coste económico y de recursos humanos que se tenían que emplear solo para llevar a cabo este fin. Aunque al principio existían múltiples sistemas operativos y protocolos se trabajó en un sistema que estandarizara la los procesos. La industria del automóvil que intercambiaba diariamente cantidades inmensas de datos para que su cadena de suministros funcionase con normalidad creo el “Comité Coordinador de los Datos de Transporte” (TDCC), a través de esto el instituto Estadounidense de Normalización creo el “estándar de mensajería x12” para sectores comerciales y no comerciales. Esto daría la posibilidad a las empresas de poder realizar transacciones y el intercambio de información de carácter comercial. (EDI Academy, s.f.)

Cuando se crearon las primeras computadoras en la década de 1970 comenzaron las relaciones comerciales a través de ellos. Pero estos primeros dispositivos no podían ofrecer un buen servicio, eran muy limitados.

En el año 1980 el comercio electrónico sufre una revolución con la ayuda de la televisión, esos catálogos que se usaban desde los años 20 ahora se podían ver a través del televisor, y no era un dibujo o una fotografía sino una imagen en movimiento. El producto podía ser mostrado en todo su esplendor, explicándolo a fondo, mostrando sus características y ofreciendo el mismo realismo como si de la misma tienda se tratara. El comprador veía el producto y llamaba por teléfono dando sus datos bancarios a través de donde compraba.

En el año 1989 se vuelve a dar un salto exponencial tecnológico con la creación de la World Wide Web o WWW. Creada por el británico Tim Berners-Lee con la ayuda del belga Robert Cailliau. Este nuevo sistema de comunicación entre computadoras cambiaria completamente los parámetros de la vida tecnológica y de la forma de comunicarse. Esta nueva forma de comunicación entre computadoras consiste en un sistema de ficheros con documentos de hipertextos vinculados entre sí en internet accesibles a través de un programa de navegación web. (Delgado, s.f.)

Con esta nueva tecnología, empezó a generarse todo el comercio electrónico casi tal y como lo conocemos hoy en día, pero limitado por su época. Se empiezan a crear sitios web exclusivamente reservados al comercio en línea. De esta forma se realizaría la

primera venta completamente online. Aquí encontramos dos productos entre los cuales se cuentan que fueron los primeros. Una pizza de champiñones con extra de queso, que fue comprada a Pizza Hut en 1994 de tal forma que se eligió, se procesó y se pagó completamente online. Y alguien que compro un CD de Sting, el cantante, el 11 de agosto de 1994 a través de la web de comercio minorista americana NetMarket. Aunque las diversas fuentes no se ponen de acuerdo en cuanto a cuál fue la primera. Sin embargo, otra leyenda urbana anterior a la creación de la World Wide Web nos cuenta que la primera venta de comercio electrónico fue una bolsa de marihuana entre los estudiantes de Stanford y los del MIT utilizando Arpanet en el año 1970. (Schreiber, 2016)

Se crean también en 1995 Amazon el rey del comercio electrónico y eBay, del que hablaremos en el siguiente punto. Y dos años más tarde la empresa informática Dell fue la primera en poder registrar un millón de dólares en ingresos exclusivamente online.

A partir de estos años el comercio electrónico sufre un crecimiento exponencial que nos lleva hasta nuestros días. (Lynkoo Love Commerce, 2012)

En este momento no existen problemas velocidad, de almacenamiento, aunque todavía existen algunos de seguridad o así al menos lo percibe el público en general. Ahora empresas que llevan toda una vida en este sector tienen, que competir casi con cualquier empresa o tienda que quiera entrar en la red por lo accesible de este medio y la cantidad de herramientas y soluciones que el medio propone. Además, no supone invertir una gran cantidad de recursos como había que hacerlo en sus inicios.

Estas herramientas que propone internet pueden ser realmente sencillas ya que permite la venta a través de sus páginas, como es el caso de Aliexpress con la venta de minoristas en su pagina web, y por un coste económico muy asumible. Pero si la creación de estas paginas de forma propia pueden ser un poco mas costosas debido a su mantenimiento. (Gariboldi, 1999)

2.2 Empresas pioneras.

Las dos grandes empresas pioneras que iniciaron la revolución en este sector fueron Amazon y eBay.

Hablar de Amazon es hablar de una de las compañías de comercio electrónico más grandes del mundo y con más importancia a nivel internacional. Con la que solo la otra gigante eBay puede intentar hacerle sombra.

Amazon S.L. (sociedad limitada) es una compañía de comercio electrónico y de servicios en la nube de origen estadounidense. Su sede está situada en la ciudad de Seattle (Washington). Actualmente es la compañía de venta al por menor con más ingresos del mundo.

En el año 1994 Jeff Bezos crea la compañía después de dejar la vicepresidencia de D.E Shaw & Co. Se muda a Seattle y allí comenzó su plan de negocios especializado en internet. Como experto de Wall Street pudo comprobar como la proyección de crecimiento de internet iba a ser de más de un 2000% anualmente.

Con la idea ya en su cabeza de crear una web de comercio electrónico, pensó una lista de objetos que creyó que se vendería con mucha rapidez. Su lista inicial fue de unos 20 productos, aunque finalmente la redujo a 5. Siguió estudiando y trabajando hasta que decidió que el mejor modelo para su negocio era la venta de libros. La demanda de estos era enorme y constante a nivel mundial. Los libros tenían un precio mucho más bajo que los que vendían en las librerías por el modelo de negocio y además podían encontrar infinidad de títulos.

Mucho antes de recibir el nombre de Amazon, la empresa paso por otros. Comenzó por “Cadabra” en referencia a las palabras mágicas, pero tras que un abogado dijera “cadáver” en lugar del nombre, decidieron que lo mejor era cambiar este. Adquiere el dominio de “Relentless.com” y durante aproximadamente un año utilizan ese nombre. Aunque el nombre no era demasiado llamativo y no termino de convencerles. Finalmente llegaron al nombre “Amazon”.

La selva de las Amazonas era un lugar muy diferente y exótico en comparación de lo conocido por el ser humano. Quería que su tienda online tuviera además el mismo tamaño que el río Amazonas

Aunque durante los primeros años solo se dedicaron a la venta de libros, posteriormente se empieza a vender más artículos, comienzan la diversificación. En el año 1998 se introducen en el mundo de los discos de música y videos y en 1999 su tienda

ya se parece mucho a la actual cuando introducen videojuegos, electrónica, software, juguetería y objetos para el hogar.

En ese mismo año, la prestigiosa revista Time reconoce Jeff Bezos como “Person of the year” título muy importante que se le ha dado a personas como John F. Kennedy, Gandhi, Stalin. Lo que hacía que se viera la gran importancia de la empresa a nivel mundial y los logros sin precedentes que estos habían conseguidos.

Su catálogo ha seguido creciendo a lo largo de los años además de a ver conseguido introducirse en multitud de países. La creación de sus propios productos y ahora incluso el pensar introducir sus propios satélites dar cobertura en lugares remotos. (Mateos, s.f.)

El otro gigante de las ventas de comercio electrónico es eBay. Esta compañía creada por Pierre Omydiar, se funda para ser un sitio de subastas por internet. A la misma vez en ser una pionera en transacciones en la red, también se convierte en una pionera en subastas en red.

eBay se funda en 1995 en San José California no todavía con ese nombre si no con el Auctionweb y con esa premisa de subastas online, vende su primer artículo. Este artículo subastado y vendido por 14 \$ fue un puntero laser. En menos de un año la página tenía 40.000 usuarios las ventas y servicios ya alcanzaban cifras de altos vuelos llegando a unos beneficios de unos 6 millones de dólares de la época y en su tienda ya se había alcanzado la venta de un millón de productos.

Oficialmente se cambia el nombre del servicio de Auctionweb a eBay en septiembre de 1997.

En el año 1998, con más de 500.000 usuarios, se plantea la idea de salir a bolsa cuando su ahora ex presidenta Meg Whitman entra en la empresa. Además, se empiezan a plantear el uso de herramientas de personalización al usuario. Para ello se adquieren las empresas Jump Inc. y Up4Sale. Finalmente, uno de los gigantes del e-commerce marcaba su salida a bolsa el 4 de septiembre del año 1998.

Una de las facetas más importantes de la compañía ha sido la compra estratégica de empresas en las que han visto un gran potencial de futuro.

En el año 2002, tras observar que la mayoría de sus clientes utilizan el sistema de PayPal como forma de pago, la empresa se interesa en adquirirla. Para muchos usuarios

todavía la compra por internet conllevaba un riesgo de ser estafados o que sus datos bancarios fueran utilizados con fines ilícitos. Por eso PayPal ofrecía un sistema completamente seguro con el objetivo del comercio online. eBay vio esto y además pensó que la empresa que consiguiera generalizar un método estándar de pago podría tener una posición dentro del comercio electrónico. De esta forma finalmente decide adquirirla haciendo un intercambio de acciones, la suma del acuerdo sería unos 1.500 millones de \$

En 2005 vuelven a comprar otra empresa estratégica, Skype. El objetivo principal fue mejorar las relaciones y los canales de comunicación entre los usuarios, tanto los que compraban como los que vendían.

A pesar de que todo parecía bueno, se empiezan a detectar la venta de una gran cantidad de productos falsificados. Lo que lleva al tribunal de comercio de París a condenar a la marca en el año 2008 a pagar 40 millones de euros en favor de algunas filiales de la compañía de lujo LVMH (Louis Vuitton Moët Hennessy). Además de eso proseguirían con una multa de 50.000€ diarios si no cesaban inmediatamente con las subastas de este tipo de productos. Esta fama continuaría persiguiéndoles pues en 2010 cuando se les nombro como la página más falsificada de todo el mundo. (EFE, 2008)

La compañía a pesar de que intenta revisar lo que cada usuario venda de forma que no sea ilícito el contenido, pero es casi imposible revisar todo ese contenido.

En 2013 adquieren Baintree que hasta entonces había sido su gran competidora en este sector. Y dos años más tarde en 2015 se desase definitivamente de PayPal

Actualmente eBay puede presumir de generar ingresos de 12 mil millones de dólares anuales. Cuenta con un número estimado de 179 millones de usuarios activos y se presenta como una gran alternativa para potenciar las pymes (pequeñas y medianas empresas) ayudándolas con iniciativas para aumentar su volumen de negocio. (Lupiañez, 2020)

2.3 El comercio electrónico en la actualidad.

La gran revolución del comercio electrónico a ayudado a grandes empresas y a pymes para aumentar su crecimiento, al igual como hemos comentado anteriormente que empresas como eBay ayudan a impulsar estos negocios.

Esta revolución se debe a que cada vez más cualquier tipo de consumo, desde ocio, deporte, alimentación, viajes, se dirige a la comodidad del usuario y a la personalización de la experiencia de compra.

¿Pero cómo se está desarrollando en números?

El comercio electrónico movió un total de 2,3 billones de dólares en ventas en el año 2017. Y a pesar de la recesión económica provocada por el Covid-19, se espera que esta cifra se duplique para el año 2021

China se coloca como el país que más mercado de comercio electrónico tiene con más de 670 mil millones de dólares. Una cifra muy por encima del segundo puesto, en este caso Estados Unidos con 340 mil millones dólares de valor. Esto se debe al creciente consumo masivo de productos de china, que a pesar de que a veces su calidad deja mucho que desear y de los largos plazos de envío de sus productos, pero los precios están muy muy por debajo de la mayoría de la competencia. Páginas como AliExpress y Wish, enfocadas como venta al por menor, han conseguido un volumen de negocio digno de admirar y más teniendo en cuenta que casi todos sus vendedores son pymes.

Interesante también es observar cuales son los artículos que mas se vendieron durante el año anterior en los que podemos destacar:

- Artículos de Moda: Este es el sector más vendido. Pero también es el sector que más ha evolucionado respecto a sus políticas de devolución. El comprar unos zapatos o una camiseta y que el resultado no fuera el deseado frustraba mucho al comprador. Pero el saber que no recuperarían su dinero y que la odisea de intentarlo les costaría mas esfuerzo que ir a la tienda a comprar el artículo en persona hacia que muchos usuarios se negaran en rotundo a adquirir este tipo de artículos de moda de esta forma. En este momento las devoluciones son gratis en casi todas las tiendas además dar muchísimas facilidades a los compradores para descambiarlos por otra talla u otro producto de su agrado
- Artículos relacionador con los viajes: Cada vez mas personas prefieren contratar sus viajes y experiencias por su cuenta sin asistir a una agencia de viajes. Eliminas el intermediario y el precio disminuye además de que la mayoría de comparadoras te dan una gran seguridad. Los viajes en avión son cada vez más fáciles tramitarlos de forma online y no tener que esperar al mismo día de embarque para hacer cosas tan engorrosas como la facturación. E incluso casi cualquier medio de transporte

se puede tramitar de forma online sin necesidad de ir a una taquilla a comprar el billete.

- Tecnología en general: A pesar de el gran crecimiento que este sector ha tenido en internet las garantías de los productos siguen siendo uno de los grandes problemas. El comprarlo en una tienda física y saber que alguien va a estar en ese momento para repararlo, es lo que todavía más buscan los clientes. De forma online es mas complicado y los tramites se podrían alargar bastante. Y para esta generación tan poco acostumbrada a esperar y muy impaciente, hace que sea poco común la compra de electrónica vía online. También depende de la oferta que encontremos, dicho sea de paso.
- Productos de segunda mano: Y de nuevo aquí tendríamos que hablar de empresas como Amazon y eBay. A pesar de que mucho de su comercio sea de distribuidoras oficiales. Pero sobre todo eBay conserva aún unas grandes ventas de productos de compra y venta entre usuarios. La capacidad de poder negociar con el vendedor y que se vendan productos muy bien conservados y de calidad hace que todavía muchas personas prefieran este tipo de compra low-cost antes que comprar productos en China de fabricación barata. Ante la crecida de este sector, se comenta que muchas de estos intercambios entre usuarios ya tendrían que estar regulados y sujetos a unos impuestos, pero todavía queda un largo camino a ese punto.
- Música y libros: A pesar de la gran revolución de los formatos electrónicos como Kindle y de las descargas de música en red, todavía muchos compradores siguen guardando como pequeños tesoros sus libros o sus discos. Y estos están teniendo un gran peso en la venta online. A pesar de que una gran cantidad de compradores siguen prefiriendo ir a su librería o sus grandes almacenes para hacerse con este tipo de producto. Pero la gran mayoría prefiere la comodidad de poder comparar precios online y disfrutar su libro en casa. En cuanto a la venta de música quizá la venta de discos no sea ya tan común, pero si la utilización de aplicaciones de pago como Spotify. Aplicaciones parecidas a las plataformas digitales de series y películas. (Diaz)

Mención especial tendríamos que hacer a las plataformas de streaming que en cuestión de unos tres años han conseguido unas ganancias considerables. A pesar de no comprar ningún producto físicamente, pero si adquieres o compras una suscripción

mensual con estas páginas web que te dan acceso a todo su catálogo de cine y series. Actualmente las líderes absolutas de este sector son Netflix y HBO, aunque la entrada en este sector de Amazon Prime video y más importante aún del gigante del entretenimiento para toda la familia, Disney plus, que ofrece ya en su catálogo material tan importante como la saga Star Wars, Marvel, o incluso los Simpson, hace que la competencia y la contratación de estos servicios vaya a seguir aumentando exponencialmente durante los próximos años.

3. ESTADISTICA COMPUTACIONAL

Aquí conoceremos cómo funciona la unión entre la estadística y las computadoras y que sistemas o aplicaciones informáticas son las más utilizadas por los investigadores.

3.1 Definición y concepto

Conocemos la estadística computacional como aquellos procesos de naturaleza estadística que se realizan mediante la utilización de programas informáticos.

La estadística en si es la ciencia que estudia la recolección, análisis e interpretación de datos. Durante siglos estos datos tenían que ser manejados de manera manual y siempre teniendo la limitación de la equivocación con el manejo de grandes cantidades de datos. A partir del descubrimiento de los ordenadores y su continua mejora, se comienza a pasar muchos de estos procesos estadísticos a través de ordenadores. Estos son capaces de procesar cantidades ingentes de datos y de tener unos márgenes de errores realmente pequeños.

Manejar estos datos también requiere de herramientas especializadas. Para cualquier tipo de investigación, la complejidad estadística es muy elevada y por eso utilizamos programas que nos ayuden a trabajar de la forma adecuada.

Existen multitud de programas estadísticos, la gran mayoría de las personas se dedican al estudio de datos o no alguna vez han manejado Microsoft Excel. Una herramienta muy versátil que permite multitud de cálculos para la cual como he comentado no es necesario un extenso conocimiento para poder usarla.

SPSS es una herramienta estadística mucho más completa, pero su licencia es de pago y es bastante costosa. Minitab que es una herramienta con menos interfaz, pero te permite tener scripts (pequeñas rutinas de programación), aunque tendría que tener conocimiento de programación para hacerlo. Statgraphics es la herramienta más sencilla de usar, pero también tiene licencia de pago.

Uno de los más utilizados, por ser un software gratuito y con una gran comunidad es el R, a continuación, lo explicare más detenidamente.

3.2 Software R Project

R es una herramienta potente de programación formada por un conjunto de aplicaciones flexibles que pueden manejarse mediante paquetes. Es un programa completamente gratuito y tiene como característica más importante ser un código abierto donde cada usuario puede crear sus propias funciones, paquetes y librerías.

El programa tuvo su primera aparición en el año 1993 de la mano del Robert Gentleman y Ross Ihaka que trabajaron intensamente desde la universidad de Auckland.

Desde el principio se mantiene la idea de que toda la comunidad de usuarios pueda ayudar y contribuir a construir un mejor entorno para R. De esta forma se consigue que la aplicación este en constante actualización y se mantenga a la par junto con los softwares de pago más importantes.

El sistema de R es muy importante para el estudio de datos y su gestión. Contando con cualquier técnica avanzada para realizar, crear gráficos de calidad en multitud de formatos, instalación en cualquier sistema operativo y además es muy fácil trabajar con datos de otros softwares de gestión de datos.

El gran problema, si se le puede llamar así de esta aplicación, es su interfaz. Al utilizar comandos en vez de una interfaz gráfica, como podría ser un programa como Skype o de su mismo entorno Spss o Excel, hace que el aprendizaje sea más lento y no para cualquier usuario. La mayoría de personas no está acostumbrada a utilizar comandos puesto que no es algo que se enseñe fácilmente o no es muy común su uso.

Aun así, existe una gran comunidad entorno al programa que ayuda conocer mejor su entorno y que ofrecen su línea de comandos para trabajar más fácilmente en cualquier

proyecto que se tenga. De esta forma una vez que se introduce en este sistema la mayoría sigue prefiriéndolo antes que cualquier otro. (Ferrero, 2018)

4. SISTEMAS DE RECOMENDACIÓN EN COMERCIO ELECTRONICO.

En este apartado veremos como son las formas de recomendación más usuales en la mayoría de páginas de comercio electrónico y nos centraremos en la forma personalización basada en los sistemas de recomendación.

4.1 Personalización en páginas webs de comercio

Si algunas de las principales páginas webs de comercio pueden destacar de otras es gracias a su sistema de personalización para mejorar la experiencia de los usuarios. Aunque es cierto que ya en casi todas las páginas web cobra un papel muy importante. Esta personalización añade muchos componentes para la experiencia de compra de una persona.

Algunas características de personalización se mantienen igual que en el comercio clásico. Ofertas personalizadas y rebajas hechas para el usuario. Ofertas de fidelización personalizada. Servicio post venta.

En el comercio electrónico se destacan técnicas que en formato tradicional de venta física no se pueden desarrollar. Técnicas que se aplican con el principio de conocer mejor a tus clientes potenciales y de fidelizarlo de modo que se continúe comprando y consumiendo, hasta el punto de llegar a crear una relación o feedback donde la empresa conozca a su cliente y a su vez le ayude a conocer a ese sector de la población. Y a su vez el cliente sienta que esa web donde compra es un sitio de confianza y que no va a quedar abandonado ante la posibilidad de recibir un mal producto. (Folguera, 2014)

Una de esas técnicas de personalización más usadas por todas las páginas webs son los sistemas de recomendación, tema central de estudio de este trabajo de fin grado.

4.2 Sistemas de recomendación

El sistema de recomendación es una herramienta que permitirá con una serie de valoraciones y datos de los usuarios poder realizar una predicción de que ayude al usuario a elegir una futura compra.

Los sistemas de recomendación actuales han avanzado y consiguen un índice de acierto final muy alto. Este funcionamiento tan mejorado de este tipo de sistemas ha podido avanzar tanto gracias al Machine Learning. Esta rama científica utiliza una inteligencia artificial y va aprendiendo de manera automática para aplicarlo a los campos que necesite.

Estas herramientas que interactúan con grandes cantidades de datos se basan en la utilización de métodos matemáticos y estadísticos ya están implementadas en casi todas las páginas webs de importancia. Ya que se ha conseguido incrementar las ventas de estos sitios webs de forma significativa. Y más actualmente son los sistemas más utilizados por las plataformas de streaming para recomendar películas o series.

Aunque se podría hablar de muchos más tipos, pero principalmente existen dos tipos de sistemas de recomendación:

- Basados en Contenido (content based filtering):

En este sistema se necesitan varios elementos. Las valoraciones de los usuarios y las características de los productos que se ofrezcan de forma cuantitativa. Esto significa que es necesario catalogar adecuadamente los productos de la lista. Pero esta información de las características no es capaz de aportar ninguna información acerca de los mismos. Las recomendaciones que se hagan se darán a razón del tipo de productos que el usuario consume y este elegirá productos con las mismas características.

- Filtrados Colaborativos (collaborative filtering)

Este tipo de filtrado funcionan usando valoraciones de los usuarios a los productos, pero sin la necesidad de conocer el producto o sus características. Solo como lo valora el usuario en sí. Este tipo de sistema es el más común por su facilidad con la que la información puede ser recopilada. Su limitación radica en el hecho de que hasta que un usuario no empieza a ver productos o puntuarlos no se le puede llegar a realizar ninguna recomendación. Los dos tipos de este filtrado son los basados en

usuarios y basado en ítems. Que son los dos filtrados que estudiaremos.
(Rodrigo, 2018)

4.3 Valoraciones y medidas de similitud

Para que los sistemas de recomendación funcionen deben de llevar unas puntuaciones o valoraciones cuantificables con los que el sistema pueda trabajar y emplear sus algoritmos. Estas valoraciones pueden medirse como por ejemplo el número de reproducciones o las veces que se ha consumido ese artículo. De forma binaria en cuanto a un si o un no, si el artículo ha sido comprado o no, etc..... Además de esto se pueden puntuar los productos de forma escalar. La forma más usual de hacer esta puntuación es de 1 a las 5 estrellas (puntos).

Los sistemas de recomendación necesitan cuantificar de alguna manera todos estos datos y de ver cuánto se parecen unos a otros. Estos parecidos se miden calculando la distancia o correlaciones que existen entre unos y otros. Los más usados son:

- Distancia Euclídea: Se define como la medida de longitud de la línea o segmento que une ambos puntos. Se calcula usando el teorema de Pitágoras.
- Distancia de Manhattan: Es la distancia entre dos puntos como el sumatorio de las diferencias absolutas entre cada dimensión.
- Distancia de coseno: Mide el ángulo que forman dos vectores y es independiente de su magnitud.
- Correlación: Existen multitud de correlaciones, Pearson, Spearman, Kendall, etc...Es una medida de distancia muy útil, ya que esta mide los patrones y no desplazamientos.

Dentro de las correlaciones, la más usada es el coeficiente de correlación de Pearson. Es muy efectivo para ámbitos muy diversos. Este índice mide el grado de covariación entre distintas variables de manera continua. Nos habla de la intensidad y de la dirección de la relación. Una de sus principales desventajas es no ser robusto cuando existen datos anómalos. Pero a pesar de todo es el más utilizado en estos sistemas de recomendación web. (Rodrigo, 2018)

5. CASO DE ESTUDIO

Ahora vamos a aplicar en un caso real como se usaría, cogiendo un conjunto real de datos. Para ello usaremos el programa R studio basado en el software del que hemos hablado R project. Y así aplicaremos los dos métodos, el filtrado colaborativo de usuarios y de ítems. Posteriormente en el siguiente punto aportaremos unas conclusiones.

5.1 Conjunto de datos de Amazon

El siguiente conjunto de datos que vamos a estudiar ha sido obtenido de la página web Kaggle donde se pueden obtener muchos más conjuntos de datos de distintas páginas webs o de cualquier tema imaginable.

El conjunto de datos en cuestión es un fichero de la empresa Amazon que cuenta con 34.660 observaciones y 21 columnas entre las que podemos encontrar el nombre del producto, la marca, el nombre del usuario y su puntuación del 1 al 5 como se hace en Amazon. (Kaggle, 2018)

Figura 1: Conjunto de datos visualizado

#	id	name	asins	brand	categories	keys
1	AVqklhwDv8e3D1O-lebb	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 16 GB - Includ...	B01AH89CN2	Amazon	Electronics;Pad & Tablets;All Tablets;Fire Tablets;Tablets,Com...	841667104676,amazon/53004484,1
2	AVqklhwDv8e3D1O-lebb	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 16 GB - Includ...	B01AH89CN2	Amazon	Electronics;Pad & Tablets;All Tablets;Fire Tablets;Tablets,Com...	841667104676,amazon/53004484,1
3	AVqklhwDv8e3D1O-lebb	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 16 GB - Includ...	B01AH89CN2	Amazon	Electronics;Pad & Tablets;All Tablets;Fire Tablets;Tablets,Com...	841667104676,amazon/53004484,1
4	AVqklhwDv8e3D1O-lebb	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 16 GB - Includ...	B01AH89CN2	Amazon	Electronics;Pad & Tablets;All Tablets;Fire Tablets;Tablets,Com...	841667104676,amazon/53004484,1
5	AVqklhwDv8e3D1O-lebb	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 16 GB - Includ...	B01AH89CN2	Amazon	Electronics;Pad & Tablets;All Tablets;Fire Tablets;Tablets,Com...	841667104676,amazon/53004484,1
6	AVqklhwDv8e3D1O-lebb	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 16 GB - Includ...	B01AH89CN2	Amazon	Electronics;Pad & Tablets;All Tablets;Fire Tablets;Tablets,Com...	841667104676,amazon/53004484,1
7	AVqklhwDv8e3D1O-lebb	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 16 GB - Includ...	B01AH89CN2	Amazon	Electronics;Pad & Tablets;All Tablets;Fire Tablets;Tablets,Com...	841667104676,amazon/53004484,1
8	AVqklhwDv8e3D1O-lebb	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 16 GB - Includ...	B01AH89CN2	Amazon	Electronics;Pad & Tablets;All Tablets;Fire Tablets;Tablets,Com...	841667104676,amazon/53004484,1
9	AVqklhwDv8e3D1O-lebb	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 16 GB - Includ...	B01AH89CN2	Amazon	Electronics;Pad & Tablets;All Tablets;Fire Tablets;Tablets,Com...	841667104676,amazon/53004484,1
10	AVqklhwDv8e3D1O-lebb	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 16 GB - Includ...	B01AH89CN2	Amazon	Electronics;Pad & Tablets;All Tablets;Fire Tablets;Tablets,Com...	841667104676,amazon/53004484,1
11	AVqklhwDv8e3D1O-lebb	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 16 GB - Includ...	B01AH89CN2	Amazon	Electronics;Pad & Tablets;All Tablets;Fire Tablets;Tablets,Com...	841667104676,amazon/53004484,1
12	AVqklhwDv8e3D1O-lebb	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 16 GB - Includ...	B01AH89CN2	Amazon	Electronics;Pad & Tablets;All Tablets;Fire Tablets;Tablets,Com...	841667104676,amazon/53004484,1
13	AVqklhwDv8e3D1O-lebb	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 16 GB - Includ...	B01AH89CN2	Amazon	Electronics;Pad & Tablets;All Tablets;Fire Tablets;Tablets,Com...	841667104676,amazon/53004484,1
14	AVqklhwDv8e3D1O-lebb	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 16 GB - Includ...	B01AH89CN2	Amazon	Electronics;Pad & Tablets;All Tablets;Fire Tablets;Tablets,Com...	841667104676,amazon/53004484,1
15	AVqklhwDv8e3D1O-lebb	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 16 GB - Includ...	B01AH89CN2	Amazon	Electronics;Pad & Tablets;All Tablets;Fire Tablets;Tablets,Com...	841667104676,amazon/53004484,1
16	AVqklhwDv8e3D1O-lebb	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 16 GB - Includ...	B01AH89CN2	Amazon	Electronics;Pad & Tablets;All Tablets;Fire Tablets;Tablets,Com...	841667104676,amazon/53004484,1
17	AVqklhwDv8e3D1O-lebb	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 16 GB - Includ...	B01AH89CN2	Amazon	Electronics;Pad & Tablets;All Tablets;Fire Tablets;Tablets,Com...	841667104676,amazon/53004484,1
18	AVqklhwDv8e3D1O-lebb	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 16 GB - Includ...	B01AH89CN2	Amazon	Electronics;Pad & Tablets;All Tablets;Fire Tablets;Tablets,Com...	841667104676,amazon/53004484,1
19	AVqklhwDv8e3D1O-lebb	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 16 GB - Includ...	B01AH89CN2	Amazon	Electronics;Pad & Tablets;All Tablets;Fire Tablets;Tablets,Com...	841667104676,amazon/53004484,1
20	AVqklhwDv8e3D1O-lebb	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 16 GB - Includ...	B01AH89CN2	Amazon	Electronics;Pad & Tablets;All Tablets;Fire Tablets;Tablets,Com...	841667104676,amazon/53004484,1
21	AVqklhwDv8e3D1O-lebb	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 16 GB - Includ...	B01AH89CN2	Amazon	Electronics;Pad & Tablets;All Tablets;Fire Tablets;Tablets,Com...	841667104676,amazon/53004484,1
22	AVqklhwDv8e3D1O-lebb	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 16 GB - Includ...	B01AH89CN2	Amazon	Electronics;Pad & Tablets;All Tablets;Fire Tablets;Tablets,Com...	841667104676,amazon/53004484,1

Fuente: <https://www.kaggle.com/>

5.2 Filtrado colaborativo basado en usuarios.

Este filtrado se basará en estudiar otros usuarios que se parezcan en cuanto a productos vistos a otro usuario, y a partir de ese punto nos hará una recomendación de 10 productos que nos puedan interesar.

Para empezar, tendremos el conjunto de datos en su versión completa, pero no todas las columnas nos interesan para realizar este estudio. Por lo tanto, tendremos que hacer un primer filtrado donde nos quedemos con las más oportunas para nuestro estudio. En este caso elegiremos ‘reviews.username’, ‘name’, ‘reviews.rating’. Y las renombraremos para su mejor uso ‘usuario’, ‘artículo’, ‘valoración’.

Ahora ya con el conjunto de datos filtrado empezaremos con el filtrado. Para esta prueba vamos a elegir a un usuario llamado ‘Mike’ y aplicamos la función de correlación para que nos diga que usuarios son los más parecido.

Figura 2: Cuadro de usuarios parecidos a ‘Mike’

usuario	similitud
1 Mike	1
2 Rick	0.989
3 Chris	0.986
4 Nick	0.984
5 John	0.981
6 Dave	0.981

Fuente: Salida de Rstudio

Como podemos observar el programa detecta que evidentemente el usuario que más se parece a ‘Mike’ pues es el propio ‘Mike’, Acto seguido nos ofrece el resto de personas de la lista con productos y valoraciones más parecidas a la suya.

El conjunto de datos, como ocurriría con otros datos, contiene muchos archivos duplicados e incluso zonas vacías donde no tenemos valores. Por lo tanto y para no tener errores en el proceso debemos seguir filtrando el conjunto. En esta ocasión lo que haremos será eliminar esos espacios vacíos y aquellos casos en los que un usuario haya votado más de una vez el mismo artículo.

Cuando vamos a pasar a detectar que productos ha visto el usuario y no, tenemos sobre todo que asumir y así escribirlo en su función, que aquel producto que aquel producto que no ha sido valorado por el usuario no ha sido visto.

El siguiente paso será predecir que valoraciones podría darle el usuario a los productos que no ha visto a través de aquellos productos de otros usuarios que se asemejan a él.

Comparando otros usuarios podemos llegar a determinar si tenemos las suficientes observaciones para un producto y en ese caso se añadirían a la lista final. De esta forma se realizaría la media ponderada de las valoraciones y así obtendríamos ya nuestras predicciones para ese usuario, volveremos a pedir 10 recomendaciones.

Figura 3: Cuadro de los productos predichos para 'Mike'

ARTICULO	predicción
1 Amazon 5W USB Official OEM Charger and Power Adapter for Fire Tablets and Kindle	0,4283
2 Amazon Fire Hd 10 Tablet, Wi-Fi, 16 Gb, Special Offers - Silver Aluminum,,,\nAmazon	0,2109
3 Amazon 9W PowerFast Official OEM USB Charger and Power Adapter for Fire Tabl	-0,0121
4 Kindle Oasis E-reader with Leather Charging Cover - Merlot, 6 High-Resolution Display	-0,0710
5 All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 32 GB - Includes Special Offers, Magenta	-0,1684
6 All-New Kindle E-reader - Black, 6 Glare-Free Touchscreen Display, Wi-Fi -	-0,9348

Fuente: Salida de Rstudio

Podemos observar como a pesar de pedir 10 recomendaciones, el programa solo nos da 6, esto se debe a que no ha podido obtener suficiente información para darnos más recomendaciones. Las observaciones y las valoraciones dadas no son suficientes para completar esa lista de 10 predicciones que le hemos hecho.

5.3 Filtrado colaborativo basado en ítems.

Este filtrado se basa en comparar los productos que el usuario ha comprado y que no ha visto, y el sistema nos ofrecerá una lista de los productos mejor valorados por el resto de los usuarios que no hemos comprado.

Para esta prueba vamos a escoger al usuario “1234” del conjunto de datos. Primero se identifica que productos ha visto el usuario y que ha valorado,

Figura 4: Cuadro productos valorados por el usuario “1234”

Amazon Kindle Paperwhite - eBook reader - 4 GB
Amazon Fire Tv,,,\nAmazon Fire Tv,,,
Amazon Fire Tv,,,\nAmazon Fire Tv,,,
All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 1
Amazon Kindle Paperwhite - eBook reader - 4 GB

Fuente: Salida de Rstudio

Se genera un grid con todas las comparaciones que se tienen que realizar. Con el objetivo de posteriormente poder revisar y que compare todos los artículos que el usuario no ha visto.

Definimos la función que calcula la similitud utilizando la correlación de Pearson. Con la función del paquete purrr , se aplica la función de correlación empleando las columnas del grid comparaciones como valores de los argumentos.

Para cada producto no valorado, se filtran los 15 productos mas parecidos y cuyo valor de similitud es mayor o igual a 0:

Figura 5: Comparación productos vistos y no vistos

Artículos no vistos	artículos vistos	similitud
"All-New Fire HD 8 Tablet, 8 HD Display, ~	"Amazon Kindle Paperwhite - eBook read~	0.577
"Amazon - Amazon Tap Portable Bluetooth a~	"Amazon Fire Tv,,,\nAmazon Fire Tv,,,"	0.200
"Amazon Fire Hd 10 Tablet, Wi-Fi, 16 Gb, ~	"Amazon Fire Tv,,,\nAmazon Fire Tv,,,"	1
"Amazon Fire Hd 10 Tablet, Wi-Fi, 16 Gb, ~	"Amazon Kindle Paperwhite - eBook read~	0.205
"Amazon Fire Hd 8 8in Tablet 16gb Black B~	"All-New Fire HD 8 Tablet, 8 HD Displa~	0.0341
"Brand New Amazon Kindle Fire 16gb 7 Ips ~	"Amazon Kindle Paperwhite - eBook read	0.310

Fuente: Salida de Rstudio

Se añade por último la valoración que el usuario 1234 ha hecho de cada uno de los productos y se calcula la media ponderada. De manera que cuando le pidamos las recomendaciones, en este caso 10 nos va a dar, nos la dará siguiendo los valores mas altos. En este caso como vamos a observar todas las valoraciones son negativas, por lo que podemos decir que, aunque se le aconsejen esas aproximadamente, pero puede que no sean del agrado del usuario.

Figura 6: Cuadro Top 10 recomendaciones al usuario ‘1234’

Top 10 recomendaciones	puntuación
1 "Amazon - Amazon Tap Portable Bluetooth and Wi-Fi Speaker - Black,, ,	-0.795
2 "Kindle Paperwhite E-reader - White, 6 High-Resolution Display (300 ppi)	-0.795
3 "vacio"	-0.824
4 "Amazon Fire Hd 10 Tablet, Wi-Fi, 16 Gb, Special Offers - Silver Alu	-1.03
5 "Fire Kids Edition Tablet, 7 Display, Wi-Fi, 16 GB, Green Kid-Proof Cas	-1.85
6 "Brand New Amazon Kindle Fire 16gb 7 Ips Display Tablet Wifi 16 Gb	-2.14
7 "All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 32 GB - Includes Speci	-2.15
8 "Kindle Oasis E-reader with Leather Charging Cover - Merlot, 6 High	-2.15
9 "Kindle Voyage E-reader, 6 High-Resolution Display (300 ppi) with Adapti	-2.15
10 "Fire Tablet, 7 Display, Wi-Fi, 8 GB - Includes Special Offers, Black"	-2.58

Fuente: Salida de Rstudio

6. CONCLUSIONES

A lo largo de todo este trabajo hemos podido estudiar desde los inicios mismos del internet, o incluso los primeros catálogos de venta por teléfono, los orígenes del comercio electrónico. Como la tecnología al igual que en otros muchos aspectos de la vida han evolucionado del comercio más clásico de nuestra tienda mas cercana a la globalización. Esta globalización a veces nos ha pasado factura como en esta crisis mundial por la pandemia o en crisis económicas anteriores, pero también nos ha acercado a otros lugares del mundo que parecían imposibles. La capacidad de poder comprar un producto en cualquier lugar del mundo con tan solo varios clics parece algo muy normal, pero se han necesitado decenas de años de desarrollo informático.

Este desarrollo a su misma vez a ayudado a las empresas a no quedarse solo en su zona de influencia y a poder ampliar sus beneficios y su nivel de empresa hasta poder ofrecer sus servicios o productos a cualquier lugar del mundo. Con todas estas mejoras cada vez es más difícil diferenciarse de la competencia. Por eso una de las medidas que las empresas están utilizando para poder diferenciarse de la competencia es en la personalización de la experiencia. Las empresas entienden que hacer que un cliente tenga una experiencia de compra única muy diferenciada del resto de clientes, le aporta un beneficio y una satisfacción y además las empresas gracias a esto ven crecer sus ventas en unos porcentajes muy altos.

Hablar de personalización no sería solo de esa experiencia de compra o de ese tratamiento en la venta y postventa. Trataría desde los mismos productos, en los cuales muchas empresas ya pueden incluso hacer un producto único del que solo vas tener esa unidad. Un claro ejemplo podría ser las tiendas de M&M's® que te permiten poner tus propios mensajes personalizado en cada gominola. Y todo pasaría por una experiencia de compra única. Dentro de esas experiencias algo donde se está trabajando más y casi todas las páginas webs cuentan con su propio sistema serían los sistemas de recomendación para a través de tus gustos, poder recomendarte otros productos. Que ha sido el eje central de este trabajo.

Uno de los grandes problemas que tiene la ejecución de estos procesos son los datos a tratar. La mayoría de estos datos tienen cantidades ingentes de observaciones y se necesitan computadoras de un gran tamaño para poder procesar estos datos y dar los resultados obtenidos. Y aunque el conjunto de datos sea menor, casi siempre viene

compuesto por cientos o miles de pequeños fallos de recogida que dificultan mucho el tratamiento de los datos.

En el estudio realizado, una gran parte del trabajo ha sido poder limpiar y filtrar los datos obtenidos, ya que el conjunto de datos sin tratar solo daba errores y no permitía obtener ningún resultado. Para una compañía la capacidad de poder obtener estos datos de los clientes de forma más limpia ahorraría mucho tiempo y trabajo a los investigadores.

Los tipos de filtrado que se han probado en el estudio han sido el filtrado basándonos en usuarios y basándonos en ítems. En general los dos dan buenos resultados, pero tienen sus problemas. El filtrado basado en ítems no trabaja desde un principio y necesita una revisión de artículos previas para poder empezar a funcionar y a recomendar productos. Y en el caso del basado en usuarios hemos visto como de 10 recomendaciones que le habíamos pedido solo nos ha podido dar 6. No teníamos suficiente información para poder aplicar el algoritmo y que nos diese todos los productos necesarios. Todo es una simbiosis y estos algoritmos de machine learning necesitan una gran base de datos para poder trabajar correctamente.

En una página web de comercio personalmente pondría varios tipos de estos filtrados como ya hacen algunas páginas importantes para cubrir todos los aspectos. No me quedaría con uno solo. Creo que ambos tienen características muy interesantes y se pueden incluso retroalimentar el uno del otro.

Hubiera sido muy interesante poder usar el filtrado basado en contenido, pero este tipo de algoritmos necesitaba de un conjunto de bases de datos especiales y especialmente preparadas. En una de ellas una lista de todos los artículos y sus características puntuadas. Y por otra parte el conjunto de datos de las valoraciones de los usuarios a los artículos vistos en cuestión. Como ya he explicado anteriormente este tipo de filtrado es el más utilizado por webs de streaming, donde cuando ves una película analizan esa película y te recomiendan más contenido de esas mismas características.

En definitiva, estos sistemas de recomendación son muy beneficiosos tanto para usuarios como para las empresas y el correcto uso de estos puede reportar un beneficio por supuesto para las empresas y también para los usuarios. Y forman parte de nuestro día a día sin ni siquiera pensar ni cómo pueden llegar a funcionar.

7. BIBLIOGRAFIA

- Delgado, P. H. (s.f.). *akus.net*. Obtenido de <https://disenowebakus.net/world-wide-web-www.php>
- Diaz, C. (s.f.). *Doofinder*. Obtenido de <https://www.doofinder.com/es/blog/productos-mas-vendidos-en-internet>
- EDI Academy. (s.f.). Obtenido de https://www.edicomgroup.com/es_CL/news/7343-edi-academy-el-origen-del-intercambio-electronico-de-datos.html
- EFE, A. (2008). La justicia francesa confirma la condena a eBay por vender artículos falsos.
- Ferrero, R. (2018). Que es R Sowftware. *maximainformacion.es*.
- Folguera, X. (2014). La Personalización: un elemento básico del e-commerce para fidelizar a nuestros clientes. *Advertis*.
- Gariboldi, G. (1999). *Comercio electronico: Conceptos y reflexiones basicas*. Buenos Aires: BID-INTAL.
- Kaggle. (2018). Amazon dataset.
- Lupiañez, F. (2020). *Zonavalue.club*. Obtenido de <https://zonavalue.com/estrategia/ebay-el-gigante-del-ecommerce>
- Lynkoo Love Commerce. (22 de 10 de 2012). Obtenido de <https://www.lynkoo.com/la-historia-del-comercio-electronico/>
- Mateos, S. M. (s.f.). Obtenido de <https://www.actualidadecommerce.com/que-es-amazon/>
- Rodrigo, J. A. (Marzo de 2018). Obtenido de https://rpubs.com/Joaquin_AR/370301
- Schreiber, T. (2016). Proceed to checkout: The unexpexted story of how ecommerce started. *Shopify Blog*.

8. ANEXOS: SCRIPTS DE R

A continuación, se adjuntan los scripts realizados en el programa de R.

8.1 Filtrado colaborativo basado en usuarios

Figura 7: Script de R del filtrado basado en usuarios

```
setwd("~/Victor")
library(readxl)
library(tidyverse)
library(recommenderlab)

dataset <- data.table::fread("1429_1.csv")
head(dataset)

dataset_final <- dataset[,c("reviews.username", "name", "reviews.rating")]
names(dataset_final) <- c("usuario", "articulo", "valoracion")
dataset_final %>% pull(valoracion) %>% median(na.rm = TRUE)
dataset_final %>% pull(valoracion) %>% mean(na.rm = TRUE)
dataset_final <- dataset_final %>%
  mutate(valoracion = scale(valoracion)) %>%
  ungroup()

#FILTRADO COLABORATIVO BASADO EN USUARIOS

# dfNames <- data.frame(usuario = unique(dataset_final$usuario), idUsuario =
# as.character(1:length(unique(dataset_final$usuario))))
#
# dataset_final <- dplyr::left_join(dataset_final, dfNames, by = "usuario")
# dataset_final <- dataset_final[,c("articulo", "idUsuario", "valoracion")]
# names(dataset_final)[2] <- "usuario"
# valoraciones_usuarios <- dataset_final %>%
#   spread(key = usuario, value = valoracion)

valoraciones_usuarios <- reshape2::dcast(dataset_final, articulo ~ usuario)
```

```

funcion_correlacion <- function(x, y){
  correlacion <- cor(x, y, method = "pearson")
  return(correlacion)
}

similitud_usuarios <- map_dbl(.x = valoraciones_usuarios[, -1],
                             .f = funcion_correlacion,
                             y = valoraciones_usuarios[, "Mike"])

similitud_usuarios <- data_frame(usuario = names(similitud_usuarios),
                                similitud = similitud_usuarios) %>%
  arrange(desc(similitud))
head(similitud_usuarios)

dataset_final[dataset_final$usuario=="", "usuario"] <- "vacio"
dataset_final[dataset_final$articulo=="", "articulo"] <- "vacio"

dataset_final$id_aux <- paste0(dataset_final$usuario, dataset_final$articulo)

duplicadosDataset <- dataset_final$id_aux[duplicated(dataset_final$id_aux)]
dataset_final_nodup <- dataset_final[!(dataset_final$id_aux %in% duplicadosDataset),]
dataset_final_nodup <- dataset_final_nodup[, -4]

articulos_vistas <- dataset_final %>%
  filter(usuario == "Mike" & !is.na(valoracion)) %>%
  pull(articulo)

articulos_no_vistas <- unique(dataset_final[!dataset_final$articulo %in% dataset_final[dataset_final$usuario == "Mike", "articulo"], "articulo"])

prediccion <- rep(NA, length(articulos_no_vistas))
articulo <- rep(NA, length(articulos_no_vistas))
n_obs_prediccion <- rep(NA, length(articulos_no_vistas))

```

```

for(i in seq_along(articulos_no_vistas)){
  # Usuarios que han visto la articulo i
  usuarios_articulo_i <- dataset_final %>%
    filter(articulo == articulos_no_vistas[i] &
           !is.na(valoracion)) %>% pull(usuario)
  if (length(usuarios_articulo_i) < 10){
    next()
  }
  # similitud es >= 0.
  top_15_usuarios <- similitud_usuarios %>%
    filter(similitud >= 0 & (usuario %in% usuarios_articulo_i)) %>%
    arrange(desc(similitud)) %>%
    head(15)
  if (nrow(top_15_usuarios) < 10){
    next()
  }
  valoraciones_top_15 <- dataset_final %>%
    filter(articulo == articulos_no_vistas[i] &
           usuario %in% top_15_usuarios$usuario)
  top_15_usuarios <- top_15_usuarios %>% left_join(valoraciones_top_15,
                                                    by = "usuario")
  prediccion[i] <- sum(top_15_usuarios$similitud * top_15_usuarios$valoracion) /
    sum(top_15_usuarios$similitud)
  articulo[i] <- articulos_no_vistas[i]
  n_obs_prediccion[i] <- nrow(top_15_usuarios)
}
top10_recomendaciones <- data.frame(articulo, prediccion, n_obs_prediccion) %>%
  arrange(desc(prediccion)) %>%
  head(10)
top10_recomendaciones

```

Fuente: Scripts recogidos desde (Rodrigo, 2018), adaptados a los nuevos datos.

8.2 Filtrado colaborativo basado en ítems

Figura 8: Scripts de R del filtrado basado en ítems.

```
dataset_final <- dataset_final_nodup
articulos_vistas <- dataset_final %>%
  filter(usuario == "1234") %>%
  pull(articulo)

articulos_no_vistas <- unique(dataset_final[!dataset_final$articulo %in% dataset_final[dataset_final$usuario == "1234", "articulo"], "articulo"])

comparaciones <- expand.grid(articulos_no_vistas, articulos_vistas,
                           stringsAsFactors = FALSE)
colnames(comparaciones) <- c("articulos_no_vistas", "articulos_vistas")

valoraciones <- reshape2::dcast(dataset_final, usuario ~ articulo)
# valoraciones <- dataset_final %>%
#   spread(key = articulo, value = valoracion, fill = NA)
class(valoraciones)
names(valoraciones)
valoraciones[1,]
class(valoraciones)

correlacion <- function(pelicula1, pelicula2, datos) {
  # Esta función calcula la correlación entre dos columnas de un dataframe.
  similitud <- cor(x = datos[, pelicula1], y = datos[, pelicula2],
                  method = "pearson", use = "na.or.complete")
  return(similitud)
}

comparaciones <- comparaciones %>%
  mutate(similitud = map2_dbl(.x = articulos_no_vistas,
                           .y = articulos_vistas,
                           .f = correlacion,
                           datos = valoraciones))

head(comparaciones)
```

```

comparaciones <- comparaciones %>% filter(similitud >= 0) %>%
  group_by(articulos_no_vistas) %>%
  top_n(n = 15, wt = similitud) %>%
  arrange(articulos_no_vistas, desc(similitud))

valoraciones_u1234 <- dataset_final %>%
  filter(usuario == "1234" & !is.na(valoracion))
comparaciones <- comparaciones %>%
  left_join(y = valoraciones_u1234,
            by = c("articulos_vistas" = "articulo"))

media_ponderada <- function(df){
  resultado <- sum(df$valoracion * df$similitud) / sum(df$similitud)
  return(resultado)
}

top10_recomendaciones <- comparaciones %>% group_by(articulos_no_vistas) %>%8.27
  nest() %>%
  mutate(prediccion = map_dbl(.x = data,
                             .f = media_ponderada)) %>%
  select(-data) %>% arrange(desc(prediccion)) %>% head(10)
top10_recomendaciones

```

Fuente: Scripts recogidos desde (Rodrigo, 2018), adaptados a los nuevos datos.