



COMILLAS
UNIVERSIDAD PONTIFICIA

ICAI

ICADE

CIHS

FACULTAD DE ADMINISTRACIÓN DE EMPRESAS

**EFECTO DE CAMPAÑAS DE
MARKETING SESGADAS EN
ALGORITMOS DE
COLLABORATIVE FILTERING
USADOS EN SISTEMAS DE
RECOMENDACIÓN**

Autor: María Torre de Silva Fuentes

Doble grado en Derecho y Business Analytics (E3 Analytics)

Tutor: Raúl González Fabre

Madrid
Marzo 2023

Resumen

La implementación de algoritmos de *collaborative filtering* ha ido cobrando importancia a lo largo de los años en las estrategias empresariales de las plataformas, cuya actividad se basa en los sistemas de recomendación. La originalidad de este método de perfilado consiste en que el algoritmo no pone el foco únicamente en el usuario objetivo, sino además en terceros usuarios. De esta manera, se permite la inferencia de conclusiones implícitas para proporcionar recomendaciones personalizadas, sin que el usuario tenga que especificar sus preferencias y sus gustos. Pero este modelo tiene una serie de retos y uno que no puede pasar por alto es el sesgo resultante del estudio de comportamientos humanos, que produce un detrimento estructural en el sistema de recomendación.

Este trabajo pretende ser esencialmente descriptivo y confirmar la hipótesis relacionada con el sesgo. Esto se hará mediante a una base de datos extraída de las interacciones de usuarios en la plataforma de Amazon y haciendo uso de gráficas para ofrecer una manera visual de detectar el sesgo al lector. Este último punto es de especial interés, puesto que gracias a esta representación gráfica se aporta un punto de vista real y actual sobre la materia. Por ello, la línea argumental comenzará con la explicación de la parte teórica, hasta llegar al desarrollo más práctico. Por último, se aportarán medidas preventivas para mitigar el sesgo en las bases de datos. En definitiva, este trabajo pretende subrayar la importancia del análisis de la calidad de cualquier base de datos, poniendo especial hincapié en las utilizadas para sistemas de recomendación, por sus implicaciones amplificadores.

Palabras clave: sistema de recomendación, collaborative filtering, sesgo, Amazon

Abstract

The implementation of collaborative filtering algorithms has been gaining importance over the years in the business strategies of platforms whose activity is based on recommender systems. The originality of this profiling method is that the algorithm does not only focus on the target user, but also on third party ones. In this way, it allows the inference of implicit conclusions in order to provide personalized recommendations, without the user having to specify his preferences and personal interests. But this model has a number of challenges and one that cannot be overlooked is the bias resulting from the study of human behaviors, which produces a structural detriment in the algorithm of the recommender system.

This paper aims to be essentially descriptive and to confirm the previously mentioned hypothesis related to bias. This will be done by means of a database extracted from user interactions on the Amazon platform and by making extensive use of graphs to provide a visual way of detecting bias to the reader. This last point is of special interest, since a real and current point of view on the subject is provided through this graphic representation. Therefore, the line of argument will begin with the explanation of the theoretical part, until reaching the more practical aspects. Finally, preventive measures will be provided as to mitigate the bias in the databases. In short, this paper aims to underline the importance of analyzing the quality of any database, with special emphasis on those used for recommender systems, due to their amplifying implications.

Key words: recommendation system, collaborative filtering, bias, Amazon

ÍNDICE

I.	INTRODUCCIÓN.....	- 5 -
a.	Contexto	- 5 -
b.	Objetivos.....	- 7 -
c.	Metodología.....	- 8 -
d.	Desarrollo	- 9 -
II.	SESGO EN LOS DATOS Y SISTEMAS DE RECOMENDACIÓN.....	- 10 -
a.	Concepto.....	- 10 -
b.	Tipos.....	- 12 -
i.	Sesgo de exposición	- 12 -
ii.	Sesgo de conformidad	- 15 -
iii.	Sesgo de selección.....	- 16 -
iv.	Sesgo de posición y popularidad.....	- 18 -
c.	Impacto y amplificación en el sistema de recomendación	- 19 -
III.	ALGORITMO DE RECOMENDACIÓN DE AMAZON.....	- 20 -
a.	Historia e innovación	- 20 -
b.	Descripción del algoritmo de recomendación de Amazon.....	- 22 -
IV.	ANÁLISIS DE LOS DATOS.....	- 25 -
a.	¿Existe el sesgo?	- 27 -
i.	Proceso de selección del producto.....	- 28 -
ii.	Nivel de satisfacción del producto	- 31 -
V.	MITIGACIÓN DEL SESGO.....	- 36 -
VI.	CALIDAD DE LA BASE DE DATOS.....	- 38 -
a.	Concepto.....	- 38 -
b.	Soluciones.....	- 39 -
VII.	CONCLUSIÓN.....	- 40 -
VIII.	BIBLIOGRAFÍA.....	- 43 -
IX.	CÓDIGO UTILIZADO	- 46 -

ÍNDICE DE FIGURAS

I.	INTRODUCCIÓN.....	- 5 -
II.	SESGO EN LOS DATOS Y SISTEMAS DE RECOMENDACIÓN.....	- 10 -
III.	ALGORITMO DE RECOMENDACIÓN DE AMAZON.....	- 20 -
IV.	ANÁLISIS DE LOS DATOS.....	- 25 -
	<i>FIGURA 1: Tabla descriptiva de la base de datos.....</i>	<i>- 27 -</i>
	<i>FIGURA 2: Tabla descriptiva de base de datos “limpia”.....</i>	<i>- 27 -</i>
	<i>FIGURA 3: Test de Chi-Cuadrado.....</i>	<i>- 28 -</i>
	<i>FIGURA 4: Tablas de contingencia y proporción (user_attr y model_attr).....</i>	<i>- 29 -</i>
	<i>FIGURA 5: Heatmap de la tabla de contingencia.....</i>	<i>- 30 -</i>
	<i>FIGURA 6: Gradiente de la tabla de contingencia.....</i>	<i>- 30 -</i>
	<i>FIGURA 7: Tablas de contingencia y proporción (user_attr y rating).....</i>	<i>- 31 -</i>
	<i>FIGURA 8: Histograma de frecuencia sobre los ratings.....</i>	<i>- 31 -</i>
	<i>FIGURA 9: Histograma comparativo de los ratings de cada categoría de model_attr.....</i>	<i>- 32 -</i>
	<i>FIGURA 10: XYPlot comparativo de los ratings de cada categoría de user_attr.....</i>	<i>- 32 -</i>
	<i>FIGURA 11: Comparativa gráfica de la media en los ratings de cada categoría de user_attr.....</i>	<i>- 33 -</i>
	<i>FIGURA 12: Comparativa gráfica de la media en los ratings de cada categoría de model_attr.....</i>	<i>- 33 -</i>
	<i>FIGURA 13: Comparativa gráfica de los ratings de cada categoría de producto.....</i>	<i>- 34 -</i>
	<i>FIGURA 14: Comparativa numérica de la media en los ratings de cada categoría de user_attr.....</i>	<i>- 35 -</i>
	<i>FIGURA 15: Comparativa numérica de la media en los ratings de cada categoría de model_attr.....</i>	<i>- 35 -</i>
	<i>FIGURA 16: Comparativa numérica de la media en los ratings de cada producto.....</i>	<i>- 35 -</i>
V.	MITIGACIÓN DEL SESGO.....	- 36 -
VI.	CALIDAD DE LA BASE DE DATOS.....	- 38 -
VII.	CONCLUSIÓN.....	- 40 -
VIII.	BIBLIOGRAFÍA.....	- 43 -
IX.	CÓDIGO UTILIZADO.....	- 46 -

I. INTRODUCCIÓN

a. Contexto

Durante generaciones, la población ha llevado desarrollado el negocio de la compraventa de bienes. Esto consiste en la transacción entre dos partes con el fin de satisfacer las necesidades de ambas, las materiales del comprador y las financieras del vendedor. A lo largo de la historia, se han planteado diferentes maneras de llevar a cabo este acuerdo para mejorar el rendimiento. Al principio, se preocupaban del artículo que intentaban vender y de la publicidad de sus características para atraer a clientes potenciales. Sin embargo, poco a poco, las campañas publicitarias fueron poniendo el foco en el consumidor y sus necesidades. En 1954, el filósofo Drucker (1954) empieza a introducir la idea cuando dice lo siguiente: “*Es el cliente quien determina lo que es una empresa, lo que produce y si va a prosperar*”.

Los sistemas de recomendación se han servido de factores como la aproximación al consumidor o el acceso a herramientas tecnológicas para desarrollar los algoritmos que utilizan. De esta manera, las empresas son capaces de llevar a cabo un filtrado de clientes y ofrecer productos personalizados, mediante procesos de perfilado basados en la interacción del consumidor con los productos disponibles. El análisis de esta valiosa fuente de información es fundamental para la posterior oferta de un variado rango de posibles productos que podrían captar la atención del consumidor. El valor de este proceso se encuentra no solo en el aumento de ventas experimentado por el vendedor, sino también en la experiencia del comprador que se ve mejorada exponencialmente en términos de eficacia y precisión.

De entre todas las técnicas utilizadas en los sistemas de recomendación, una de las más frecuentes es la de *collaborative filtering*. El algoritmo funciona a raíz de una comparativa paralela tanto del usuario como de las ofertas de una misma plataforma. Es una manera singular de llevar a cabo el objetivo por dos razones, principalmente: el uso de las características correspondientes a otros usuarios y las conclusiones implícitas. En primer lugar, el sistema no focaliza sus observaciones únicamente en el usuario objetivo, sino que se centra además en terceros usuarios para llevar a cabo las recomendaciones apropiadas. Por lo tanto, no se pone el punto de mira únicamente en ese usuario objetivo, sino que se utiliza información sobre otros similares a él para aportar un punto de vista enriquecido sobre sus preferencias. En segundo lugar, el sistema puede inferir de manera implícita las prioridades de cada cliente, sin necesidad de que el usuario señale

explícitamente sus gustos. Así, un usuario que elija un producto, será automáticamente comparado con el resto de usuarios para clasificarlo en una categoría con características similares y se le recomendará algo acorde a ese segmento. Un ejemplo a través del cual vemos representado claramente estas dos características del sistema de recomendación se puede apreciar en empresas como Netflix, que recomienda películas que coincidan tanto con la categoría especificada por el propio cliente, como con sus previas visualizaciones. Por ejemplo, al cliente se le recomendará la película de “*Top Gun*” por haber elegido explícitamente la categoría de películas de acción y, por otro lado, por haber visto “*Encanto*” se le clasificará en el segmento de niños y se le recomendará una película popular entre las personas que encajan en ese perfil: “*Frozen*”.

En definitiva, el sistema hace uso de *Machine Learning* para aprender a recomendar de manera acertada y personalizada a cada cliente a base de la reiteración constante y la búsqueda de patrones en los datos que se le presentan. Este bucle interminable es llevado a cabo por un programa tecnológico, por lo que sería fácil pensar que no existe ningún tipo de sesgo humano en él. Pero el hecho de que el sistema esté basado en datos recogidos de manera automática, no lo hace más neutral. La parcialidad humana acaba por interferir en el proceso tecnológico de una manera u otra puesto que los datos estudian precisamente comportamientos humanos.

En línea con lo anterior, aunque el sistema del que hablamos parece ser muy exacto, el estudio de las mencionadas interacciones puede entrañar problemas en caso de que las interacciones resulten estar sesgadas por la forma en la que se comercializa el producto. Por ejemplo, la selección de un modelo humano concreto u otro en la imagen que se utiliza para la venta del producto podrá tener implicaciones muy distintas y afectar al resultado de las interacciones que debe estudiar más tarde el sistema. Por lo tanto, el mismo producto se puede llegar a presentar de maneras muy distintas, mediante un rango muy amplio de imágenes que estén dirigidas específicamente a determinados géneros, edades o grupos raciales, entre otras categorías. Hay que ser extremadamente cuidadosos en la determinación de las estrategias mencionadas puesto que el impacto en el comportamiento de los usuarios puede ser enorme. La correlación entre la estrategia utilizada y la elección del usuario puede llegar a dar lugar a una infrarrepresentación de segmentos dentro del mercado en las interacciones del consumidor. Por ejemplo, un usuario femenino al que le puedan atraer los productos relacionados con el motociclismo puede llegar a ser menos propenso a interactuar con ellos si se promocionan utilizando

imágenes que no estén dirigidas a su nicho de mercado, es decir, imágenes consideradas más masculinas.

El problema del sesgo en las bases de datos cuando hablamos de algoritmos de recomendación, ha sido un tema de interés los últimos años por ser uno de los sistemas más comunes en grandes plataformas como Amazon, Netflix o Spotify. Así, autores como Baeza-Yates (Peiró, K. & Baeza-Yates, R., 2021) han hecho una clasificación de los distintos sesgos, para abordar la cuestión de la manera más eficaz posible. Los tres principales grupos en los que se han definido los sesgos son: (i) estadístico, que consiste en la desviación consistente y significativa de una distribución previa; (ii) cultural, que incluye todas las opiniones y valoraciones adquiridas a lo largo de nuestras vidas; (iii) cognitivo, se trata de una variación sistemática de la norma. El sesgo que nos concierne para la explicación de este trabajo es un tipo de sesgo cognitivo, el sesgo de marketing. Este sesgo describe una tendencia irracional en el proceso de decisión del cliente y es reflejo de los factores que afectan a las actitudes de los clientes respecto a los productos que se les presentan. Aun así, existen numerosas y diversas clasificaciones en lo que se refiere al sesgo en los datasets que desarrollaremos más adelante.

Esto supone una potencial pérdida para los vendedores, una experiencia insatisfactoria por parte del usuario y, por último, pero no menos importante, supone un preocupante problema social.

b. Objetivos

La idea principal que se quiere extraer de la base de datos con la que contamos es identificar y describir sesgos de marketing en los datos de interacción consumidor-producto y el efecto que tiene en los algoritmos de recomendación. El objetivo es exponer el hecho de que las estrategias de marketing que ponen el foco en específicos grupos de personas producen un set de datos que, por resultar sesgado, supone un detrimento del sistema de recomendación por parte de plataformas de venta como Amazon. Un sistema de recomendación que se ve comprometido por una base de datos sesgada tiene un número de consecuencias perjudiciales desde la dificultad de los usuarios para encontrar productos relevantes, hasta problemas éticos y sociales.

Para conseguir el objetivo principal previamente mencionado nos centraremos en comprobar dos puntos. Para empezar, si efectivamente este sesgo de marketing existe en bases de datos de comercio electrónico en mundo real. Una vez verifiquemos lo anterior,

el siguiente paso consistirá en explicar cómo esta parcialidad puede verse amplificada a lo largo del bucle del sistema de recomendación, así como las soluciones posibles para abordar el problema.

c. Metodología

A la hora de conseguir exponer el impacto de las bases de datos sesgadas en los sistemas de recomendación, llevaremos a cabo un procedimiento formado por dos partes. Para empezar, se establecerá el contexto y se llevará a cabo un desarrollo teórico. Esta primera parte versará acerca de los sistemas de recomendación y sus diferentes dificultades, cerrando el foco específicamente sobre los sistemas de recomendación utilizados por Amazon y las dificultades que estos entrañan.

Más adelante, se abordará la parte práctica del asunto. En ella, analizaremos una base de datos elegida para ilustrar el tema a tratar y se expondrán las conclusiones de ese análisis. Después de haber entendido la base de datos que proponemos para exponer el tema, el análisis se realizará siguiendo los siguientes puntos.

Para empezar, se estudiará el proceso de selección del producto. En el dataset escogido se asocian una serie de atributos a cada uno de los productos. Para llevar a cabo el estudio pertinente, se examinará la existencia o no de una correlación entre el atributo que se asocia a cada producto y la identidad del usuario en términos de frecuencia de interacción en el conjunto de datos. Para realizar este análisis utilizaremos la Prueba de Chi Cuadrada de Pearson. Ésta puede ser usada para comprobar la asociación dos variables en términos de frecuencia. Para diferenciar aún más el posible sesgo de marketing de los efectos de tendencia, realizaremos otras pruebas de asociación en los datos de interacción completos, así como en las interacciones dentro de diferentes intervalos de tiempo. Una vez obtenido el P-valor de la independencia de las variables, podremos llegar a conclusiones claras sobre el tema. Además, se presentará una tabla de contingencia de la distribución de frecuencia entre los atributos de cada uno de los productos y la identidad del usuario.

En segundo lugar, se estudiará el nivel de satisfacción con el producto que ha tenido ese cliente. Para ello se utilizarán las mismas variables que en el punto anterior (variables de atributos del producto e identidad del usuario), así como los ratings que ha decidido dejar el usuario después de su experiencia en la compra del producto. Para

conseguir esto, se utilizarán pruebas de asociación entre el atributo del producto y la identidad del usuario, pero, esta vez, poniendo especial atención en los ratings.

Por último, ambas partes del análisis contarán con sus respectivas ilustraciones gráficas que representen las relaciones que nos interesan para una mejor comprensión del asunto.

d. Desarrollo

En el segundo capítulo, sesgo en los datos y adecuación de los sistemas de recomendación, se estudiarán los sesgos en los datos de manera general. Por un lado, se llevará a cabo una explicación y exposición de las distintas categorizaciones que se pueden hacer de los sesgos existentes. Para resaltar las divergencias entre los tipos de sesgo, relataremos las diferentes razones o causas por las que se puede dar esta parcialidad sobre los datos, que son las que determinan el tipo de sesgo. Por otro lado, se verá el impacto que una base de datos parcial puede tener sobre un sistema de recomendación y cómo puede verse amplificado a lo largo del bucle. Este último punto será explicado desde un plano teórico y tendrá especial importancia a la hora de presentar la propuesta de remedios a nuestro problema.

El tercer capítulo, relativo al algoritmo de recomendación de Amazon, servirá como lugar para hablar del contexto histórico y la descripción a fondo del sistema de recomendación de Amazon. Esta sección servirá para determinar las peculiaridades del sistema utilizado por esta única e innovadora plataforma de compraventa. Además, es importante para enmarcar la base de datos que describiremos en uno de los siguientes puntos.

El cuarto capítulo de nuestro índice es a la que habrá que prestar especial atención por su carácter analítico y explicativo. Primero, nos centraremos en la base de datos: detallaremos las características y propiedades de nuestra base de datos y explicaremos las razones que la hacen interesante para nuestro análisis. Más adelante, utilizaremos la metodología explicada previamente y, después de estudiar los dos aspectos mencionados, deberíamos de ser capaces de llegar a la conclusión de que la publicidad centrada en nichos específicos del mercado da lugar a una base de datos sesgada, que distorsiona los resultados del sistema de recomendación. La identificación de la relación que existe entre la manera focalizada en la que se publicitan los productos y los sistemas de recomendación posteriores tiene especial interés a la hora de proponer soluciones.

El apartado quinto del trabajo explica la importancia del estudio sobre la calidad de los datos a la hora de usar una base de datos en un sistema de recomendación.

El apartado sexto y último habla sobre los indicadores a los que hay que prestar atención a la hora de asegurarse de mitigar el sesgo. Estas dos soluciones preventivas para luchar contra el sesgo de una base de datos son un acercamiento a la infinidad de estudios y escritos que hay sobre el tema.

Las conclusiones previas nos llevan al séptimo apartado relativo a posibles vías de mejora que puedan servir para evitar futuras parcialidades en los sistemas de recomendación. Para explicar las distintas soluciones tomaremos en cuenta no solo el impacto de las bases sesgadas sino, además, el efecto amplificador que se produce por tratarse un bucle automático.

II. SESGO EN LOS DATOS Y SISTEMAS DE RECOMENDACIÓN

a. Concepto

La eficacia de los sistemas de recomendación radica en su capacidad para procesar grandes cantidades de datos y ofrecer al usuario recomendaciones personalizadas en función de sus preferencias y comportamientos de consumo. La base de semejantes sistemas son los datos observacionales o, más bien, sobre las interacciones entre usuario y producto. De esta manera, los datos usados para que el algoritmo se entrene se ven influenciados por una serie de variables no deseadas como la opinión de los usuarios. El sesgo que generan en el modelo de recomendación este tipo de elementos acaba por dar un resultado poco verídico y realista de la elección del usuario. El reflejo se ve sesgado e influenciado por la opinión de terceros usuarios porque las recomendaciones manifiestan de forma dominante las opiniones de un sector particular de consumidores. Lo más probable es que esta distorsión de la realidad no sea una proyección representativa de la diversidad de los gustos de la población.

Por lo tanto, el sesgo en la base de datos por las opiniones de los usuarios se genera por la dominante presencia de las opiniones de un grupo de usuarios. Es preocupante en tanto en cuanto se dejen de lado las preferencias y opiniones de usuarios con gustos diferentes o minoritarios.

El hecho de que los datos sean observacionales significa que se basan en las interacciones de los usuarios. Como se ha mencionado anteriormente, debido al uso de semejantes datos, la muestra puede verse afectada por razones como las opiniones del conjunto de individuos, lo que puede hacer que la distribución de los datos de entrenamiento difiera significativamente de la distribución de los datos de prueba. El riesgo que esto plantea es que el resultado del modelo de recomendación sea poco óptimo, porque la diferencia entre la distribución de los datos de entrenamiento (*training data*) y la de los datos de prueba (*testing data*) provocaría una imagen puntual y sesgada de las preferencias de los consumidores. En consecuencia, de semejante discrepancia, el algoritmo se encontraría con la difícil tarea de pluralizar de manera exacta con unos *training data* y *test data* incoherentes lo que le llevaría a dejar de ser eficiente y disminuir en su rendimiento considerablemente. Esto es porque las interacciones aprendidas por el algoritmo en el *training data*, pueden no tener nada que ver con las que aplican al *testing data* y como resultado, el modelo puede dejar de ser de utilidad al enfrentarse a nuevos ejemplos de interacciones, para las que no ha sido entrenado y que no tienen este sesgo.

De esta manera, el sesgo de datos se puede describir como la disparidad entre lo que se espera que un modelo de recomendación produzca y lo que realmente sucede. Si existe una gran diferencia entre los valores predichos y los valores reales, esta discrepancia puede afectar negativamente la eficacia del sistema. Para detectar el sesgo, es crucial comparar la media de los valores predichos con la media de los valores reales. Normalmente, un sistema de recomendación se entrena con bases de datos antiguas, que tengan variables iguales a las que se quieren estudiar en el presente y, mediante un sistema de repetición constante aprende a proporcionar la recomendación correspondiente cuando se le aportan una serie de datos concretos (Simeone, O., 2018). De esta manera, podríamos decir que los datos predichos son aquellos que el sistema recomienda al usuario, mediante un trabajo repetitivo de datos antiguos con características similares, mientras que los datos reales son la elección efectiva del usuario en cuestión. A la hora de encontrar un sesgo mediante el contraste entre estos dos tipos de datos, habría que seguir una serie de pasos. Para empezar, una vez se identifique esta discrepancia, se tendrá que analizar la diferencia entre ellos para buscar tendencias sistemáticas que nos indiquen el sesgo. Por ejemplo, si el sistema de recomendación tiende a sobreestimar el interés que va a tener el usuario en un producto según si es de género masculino o femenino, habrá evidencias de que existe un sesgo en esa variable específica en nuestra base de datos original, a través de la cual se ha entrenado el algoritmo. Finalmente, para verificar si las

diferencias son completamente aleatorias o efectivamente existe un sesgo, se podrán realizar pruebas estadísticas que confirmen o rechacen nuestra teoría. Un ejemplo de este tipo de pruebas es la prueba de chi-cuadrado, que utilizaremos más adelante para analizar nuestro conjunto de datos. Aun así, existen más pruebas estadísticas (Moreno, L., 2023) de este estilo que se pueden utilizar como las pruebas de: t-test, z-test o ANOVA.

El sesgo no solo se encuentra en los datos sino también en los algoritmos y en la manera en la que se realiza el tratamiento de datos en concreto. El impacto que puede llegar a tener el sesgo, en la precisión y la confianza que se puede inferir de los resultados obtenidos, es desde luego considerable. Por ello, es recomendable aplicar las medidas apropiadas para intentar minimizar el impacto del sesgo de los datos. Solo así podremos mitigar la significativa influencia de este sesgo sobre la base de datos de un sistema de recomendación que favorece a una muestra limitada de usuarios, dejando indebidamente de lado a otros.

Tras dejar claro el concepto de sesgo en los datos, corresponde ahora definir los distintos tipos de sesgo que existen en cuanto a los datos. Podríamos diferenciar las desviaciones de muchas maneras, aunque la clasificación que mejor se adecúa a la descripción propuesta en el párrafo anterior sería dividirlos como sesgos de: selección, exposición, conformidad y posición.

b. Tipos

i. Sesgo de exposición

Primero, el sesgo de exposición se define como aquel que surge entre los datos de un mismo tipo como consecuencia de la exposición a un subconjunto limitado de productos. Esta exposición parcial de la variedad de ofertas existentes influye considerablemente en las preferencias del usuario y puede llegar a tener un gran impacto en las recomendaciones que el sistema le proponga en un futuro. En otras palabras, los usuarios son más propensos a interactuar con los elementos que se les presentan con mayor frecuencia, lo que produce la perpetuación de la exposición de este segmento de productos al usuario. Existen varias razones por las que se da este tipo de sesgo.

En primer lugar, (Sun, T., Zhang, Y., Hu, X., Xie, X., & Zhang, Y., 2021) los modelos que se hubieran implementado en los sistemas de recomendación previos influyen en la exposición porque a través de ellos se determina lo que mostrar.

En segundo lugar, las interacciones de los usuarios suponen un factor crítico en este sesgo. Como se ha comentado previamente, la facultad que tienen los usuarios en lo relativo a la búsqueda y elección activa y libre de los elementos de interés supone que la selección del usuario desempeñe un papel fundamental en la exposición que sufren los datos. La retroalimentación implícita se emplea con frecuencia en la recomendación, que refleja las acciones típicas de los usuarios, como los clics, las visualizaciones y las compras. A diferencia de la retroalimentación explícita, que proporciona evaluaciones numéricas, la implícita se limita a transmitir una parte de una señal positiva. Debido a la falta de conocimiento sobre lo que no le gusta al usuario, el aprendizaje debe basarse en encuentros no observados y extraer de ellos la señal negativa. En concreto, las razones por las que una interacción no quede declarada entre usuario y artículo son dos: bien el producto no acaba de ajustarse a los intereses específicos del cliente, o bien simplemente el cliente desconoce de la existencia de este producto. De esta manera, teniendo en cuenta que los productos podrían ser elegidos o no, hay interacciones que nunca llegan a tener lugar, es decir, elementos que no se recomiendan al usuario y, por lo tanto, no se llegan a elegir, a las que llamaremos interacciones no observadas. Este tipo de interacciones pueden llegar a ser engañosas, porque el hecho de que las interacciones sean “no observadas” no quiere decir sistemáticamente que la respuesta del usuario sea negativa. Puede darse, en línea con lo mencionado previamente acerca de la elección libre y activa de los usuarios, que el usuario simplemente no tenga conocimiento del elemento en cuestión. De ahí el carácter engañoso de las interacciones afectadas por el sesgo de selección y la difícil tarea que entraña la interpretación de estos datos. Así, podríamos hacer una clasificación (Chuah, G. S. F., Ngo, T. D., Pham, T. V., Lim, C. P., & Chua, T. S., 2021) de las interacciones en: (i) las que existe una selección activa (positivas); (ii) las que no hay selección (negativas); (iii) en las que no se da la selección por una falta de consciencia acerca del elemento (negativas potencialmente positivas). Teniendo en cuenta la dificultad de discernir la delgada línea que existe entre los dos últimos tipos, la interpretación de las interacciones que pasan desapercibidas resulta una tarea compleja y da lugar a sesgos pronunciados. Ambas son no observadas y parece que ambas reflejan una preferencia por lo negativo, pero hay que tener en cuenta que una de ellas no es negativa realmente.

En tercer lugar, la trayectoria personal de cada usuario y el contexto en el que vive afectan radicalmente a la exposición del mismo y a la respuesta que dará a cada elemento. Por poner un ejemplo, establezcamos un escenario en el que se esté estudiando el

comportamiento de dos usuarios. El usuario1 es un hombre de cincuenta años, padre de tres hijos y recientemente desempleado. El usuario2 es un viudo de cincuenta años, sin hijos, en la cresta de su carrera profesional. La oferta de un nuevo iPhone probablemente no recibirá la misma respuesta de cada uno de estos usuarios porque su estado social, económico y personal no son los mismos. En esta dimensión del sesgo de exposición, se podrían también considerar factores como la comunidad de la que forma parte el usuario o la ubicación geográfica.

En cuarto lugar, la popularidad de cada elemento afecta profundamente en la reacción de los usuarios, que normalmente son más proclives a interactuar con elementos populares, por la confianza que supone algo conocido, que con los elementos que no lo son tanto. Este factor lo estudiaremos con un mayor grado de profundidad más adelante, puesto que el impacto que supone a la base de datos lo merece.

Estas cuatro situaciones, que ponen al usuario en una predisposición sesgada antes de que lleve a cabo su elección, son distintas formas que unificamos (He, X., Liao, L., Zhang, H., Nie, L., Hu, X., & Chua, T. S., 2017) en el sesgo de exposición para ser concisos en nuestro análisis.

El sesgo de exposición es perjudicial porque produce un refuerzo del sesgo de la base de datos en caso de que ya lo hubiera previamente. El sistema, como resultado de este sesgo reforzado ofrece recomendaciones uniformes, que limitan considerablemente la libertad de elección del usuario. Este último, solamente se vería expuesto a una serie de productos tasados, que no son realmente todos los disponibles y puede incluso dejar desapercibidos algunos productos que podrían haber sido relevantes o atractivos para el usuario si se le hubieran ofrecido. Por si fuera poco, el refuerzo del sesgo acentúa la discriminación de usuarios o productores y da lugar a recomendaciones inexactas o erróneas, de manera que al final se perpetúan las desigualdades predispuestas. En el ejemplo propuesto anteriormente, en el que a un usuario únicamente se le ofrecían una serie de productos limitados en virtud de alguna de sus características identificativas (como la edad, la raza o el género), se verían perjudicados tanto el usuario como el vendedor del producto que no ha sido representado adecuadamente en la página web del usuario (Stahl, B. C., Schroeder, D., & Rodrigues, R., 2023). En última instancia, tanto el usuario como el vendedor de los productos acabarían por perder confianza en el sistema de recomendación claramente sesgado como resultado de este sesgo reforzado.

ii. Sesgo de conformidad

Segundo, el sesgo de conformidad es el que vemos representado en los casos en los que el usuario forma parte de un grupo y actúa de manera parecida al conjunto (Guimond, S., & Dambrun, M., 2018). Esto ocurre cuando el usuario no actúa de manera independiente y su elección no refleja sus propias preferencias, sino las del grupo en general. Este sesgo, que afecta a la independencia de las valoraciones del usuario, se suele dar sobre todo en elementos que son previamente presentados al público. De esta manera, las interacciones se ven sesgadas por la influencia que ejerce el grupo sobre el individuo y no se podría estudiar un reflejo realista de las preferencias de los usuarios. La consecuencia inmediata de que los resultados no reflejen las preferencias individuales reales de los usuarios, sino la opinión del grupo al que pertenecen, suele provocar un pronunciado impacto negativo en la precisión de los datos recopilados en estudios de mercado o encuestas de opinión. Por otro lado, teniendo en cuenta las características de este sesgo, se podría aprovechar para otros aspectos como la promoción de hábitos saludables de un grupo social (Asch, S. E., 2018). Aun así, para el caso que hoy nos ocupa, relacionado con la evaluación de productos y comportamientos de los usuarios el efecto es, como hemos mencionado previamente, claramente negativo sobre la calidad de datos recopilados.

En resumidas palabras, este sesgo nos supone un importante impedimento a tener en cuenta a la hora de estudiar y analizar los comportamientos y productos de un sistema de recomendación, ya que puede influir en la opinión de los usuarios y afectar la precisión de los datos recopilados. Habría que tomar medidas para mitigar el impacto de este sesgo sobre nuestros datos si se diera.

Los problemas principales que este sesgo plantea se pueden dividir en dos: el impacto en la personalización de los productos y el efecto de burbuja informativa (Flaxman, S., Goel, S., & Rao, J. M., 2016).

En primer lugar, el sesgo de conformidad hace que el sistema de recomendación trate a cada usuario no como un individuo independiente, sino más bien como parte de un grupo definido. Esto habilita al algoritmo a presuponer aspectos del usuario que pueden o no ser ciertas por el mero hecho de pertenecer a un grupo determinado. La focalización del sistema de recomendación en la opinión popular y la posterior aplicación de ésta al individuo, lo único que hace es empobrecer la fuente de información que recibe el usuario, viéndose así las oportunidades de encontrar el producto más adecuado para él

considerablemente reducidas. El resultado de esta reducción de oportunidades para el usuario y la aplicación de la opinión popular por tener una serie de características concretas, es que el sistema de recomendación deja de ser capaz de personalizar según el individuo y acaba por ofrecer recomendaciones poco certeras y acertadas al usuario.

En segundo lugar, este sesgo produce el fenómeno denominado burbuja informativa. Esto se da en los supuestos de hecho en los que al usuario solamente se le presenta información de un tipo determinado. El empobrecimiento de diversidad en la oferta de productos al usuario, hace que éste solo pueda acceder a información y opiniones que suelen estar alineados con su punto de vista inicial. De esta manera, las decisiones del usuario no solo estarían influenciadas sino, incluso limitadas a un rango muy específico de artículos. Las recomendaciones del algoritmo dejarían de enriquecer al usuario, para no ser más que una iteración constante de sus opiniones.

iii. Sesgo de selección

Tercero, el sesgo de selección tiene lugar con motivo de la arbitrariedad que se manifiesta en las elecciones de valoraciones de los clientes (Kumar, A., Gupta, M., Jain, V., & Varshneya, P., 2021). Así, esta falta de uniformidad hace que las evaluaciones de los individuos concretos difieran de manera consistente de la población de general, lo que da lugar a un error sistemático en el resultado. En el contexto de los sistemas de recomendación, los usuarios tienen libertad de optar por dar su valoración sobre productos específicos en función de sus características, su precio, sus recomendaciones y su afinidad con ellos, entre muchos otros factores. Esta falta de uniformidad hace que la búsqueda de un patrón se convierta en una tarea complicada a la hora de implementar el modelo. Por lo tanto, las valoraciones observadas no son una muestra para nada representativa del conjunto completo de la población. Los problemas que presentan este tipo de sesgos a la hora de estudiar una base de datos de un sistema de recomendación se pueden resumir en cuatro: (i) extremos; (ii) ausencia de representación; y (iii) representación poco fiel de características.

En primer lugar, el hecho de que los usuarios tengan la facultad de aportar valoraciones de los productos que ellos decidan aleatoriamente, hace que las opiniones que se den sean de carácter más extremo de lo normal. Así, un usuario que ha quedado contento con un producto por ser de la calidad y tipo de producto que esperaba tiene muchas menos posibilidades de escribir una recomendación que un usuario que ha

quedado sorprendido con la buena calidad del producto en cuestión. El sentimiento de urgencia que hace que un usuario con experiencia extraordinaria o pésima sea más propenso a valorar que uno con una experiencia más moderada provoca una desmesurada falta de equilibrio en las opiniones que se aportan. Por lo tanto, un usuario que se base en las evaluaciones de otros clientes en realidad se está basando en una versión distorsionada y exagerada de la realidad.

En segundo lugar, a la hora de considerar las valoraciones de otros clientes, empresas como Amazon tienen sistemas específicos para dar más peso a unas valoraciones que a otras y así calcular las calificaciones de estrellas del producto. Amazon, tiene un algoritmo especializado en clasificar a los distintos clientes según si su opinión es reciente y si su estado de compra está verificado (Amazon, s.f.). De esta manera, se puede llegar a dar una falta de representación de la opinión de clientes cuya opinión puede ser valiosa, pero que se pasan por alto por no haber comprado en Amazon. Por ejemplo, un individuo que tenga una prenda de ropa específica, aunque no la haya comprado a través de Amazon. La opinión de este individuo es de interés puesto que, como dueño de esta prenda puede hablar de la calidad del material, la comodidad o la facilidad de desgaste de la prenda. Aun así, como no la ha comprado a través de esta plataforma, a su opinión no se le dará la misma importancia que opiniones de clientes directos de Amazon. La falta de opiniones representativas acaba generando un sesgo hacia un grupo específico de personas y se puede llegar a alejar de la realidad.

En tercer y último lugar, el carácter arbitrario de las valoraciones de los usuarios del que hablábamos previamente puede afectar a la representación de características específicas del producto. Como el usuario elige las características sobre las que va a ofrecer su opinión, se pueden representar de forma extremadamente favorable o desfavorable solo ciertos aspectos del producto, dejando otros de lado, lo que genera recomendaciones sesgadas.

En línea con lo propuesto anteriormente, la calidad de los modelos y sugerencias producidos puede verse afectada significativamente por el sesgo de selección, que es un problema que surge con frecuencia en el proceso de recopilación de datos. Por eso, es fundamental reconocer este desafío y tomar medidas para reducir sus efectos. Estas medidas pueden consistir en una recogida de datos más meticulosa y exhaustiva, la aplicación de métodos de corrección del sesgo a lo largo del proceso de modelización y el uso de técnicas estadísticas sofisticadas para ajustar el sesgo.

En definitiva, la precisión y el calibre de los sistemas de recomendación, así como otros modelos de datos pueden verse aumentadas de manera considerable mediante la adopción de medidas proactivas que aborden la cuestión.

iv. Sesgo de posición y popularidad

Este tipo de sesgo se da porque hay una tendencia pronunciada a la interacción con elementos de la lista de recomendaciones del sistema que, aunque ocupen uno de los puestos superiores, realmente no necesariamente son muy relevantes en cuanto a un buen reflejo de las preferencias reales del usuario. Por ejemplo, en una lista de elementos, un usuario suele ser más propenso a examinar los primeros y descartar el resto.

El sesgo de posición es realmente parecido al sesgo de popularidad (Fleder, D., & Hosanagar, K., 2010), que se refiere a la variación de selección de productos por la popularidad que otros usuarios le confieren mediante recomendaciones. La diferencia entre ambos sesgos radica en la manera de entender la desviación entre los resultados del modelo y los resultados ideales. Mientras que el sesgo de posición se centra en los datos, es más bien un sesgo de observación y no tanto de experimentación, el sesgo de popularidad (Moliner, J., Sainz, J., Pérez-González, F., & del-Pozo-Coronado, J., 2018) se presenta en alguna de las distintas etapas del algoritmo, estudia la desviación a partir de la experimentación.

Aun así, ambos comparten las consecuencias perjudiciales para el estudio de los datos que dispongamos: el desequilibrio entre la posición de los productos en la percepción de los usuarios y sus preferencias reales. El hecho de que ambos se enfoquen en productos populares puede resultar en que otros que, quizá hubieran sido más relevantes, tengan mucha menos visibilidad. Por lo tanto, las recomendaciones pueden llegar a ser resultar uniformes y a olvidar preferencias individuales.

Adicionalmente, estos sesgos tienen un efecto exponencial (Chizari, N., Shoeibi, N., & Moreno-García, M. N., 2023) de detrimento a la parcialidad de la base de datos. Cuanto más elijan los usuarios elementos populares o con una buena posición en la lista de recomendaciones, mejor parados saldrán en las sugerencias. Por otro lado, se han estudiado una serie de soluciones factibles para evitar que el sesgo afecte a nuestras bases de datos. Entre las soluciones que se han dado destacan, sobre todo el “*re-ranking*” (Abdollahpouri, H., Burke, R., & Mobasher, B., 2019) y los análisis centrados o en los

objetos o en los usuarios (Abdollahpouri, H., Mansoury, M., Burke, R., Mobasher, B., & Malthouse, E., 2021).

Además de los sesgos presentados, no podemos olvidar el sesgo inherente a todo modelo de aprendizaje predictivo, resultado de la generalización de las interacciones con las que el modelo cuenta para llegar a predicciones coherentes. Estas suposiciones, que mejoran el aprendizaje sobre la función objetivo, son la base esencial de un sistema de recomendación, que aprende y predice en consecuencia de lo aprendido.

c. Impacto y amplificación en el sistema de recomendación

En el apartado anterior, se han entendido los sesgos que surgen de los datos, así como del propio algoritmo de los sistemas de recomendación. Sin embargo, teniendo en cuenta que los modelos de este tipo se basan en la iteración para aprender el comportamiento de las interacciones guardadas, resulta imperativo estudiar el efecto que tiene este estudio repetitivo de datos sesgados. Ya en el apartado previo se introdujo esta idea, cuando se mencionó del carácter exponencial del detrimento a la parcialidad de la base de datos, pero, ¿qué quiere decir esto?

El impacto de la repetición constante en los sistemas de recomendación es considerable porque, no solo se vería reproducido el sesgo, comentado previamente, de los datos, sino que, además, se conseguiría una amplificación de dicho sesgo. Esta amplificación (Hall, M., van der Maaten, L., Gustafson, L., Jones, M., & Adcock, A., 2022) no es más que el efecto que los datos sesgados tienen sobre el modelo, puesto que, de no existir ese sesgo, no habría ninguna desviación que reproducir (Mansoury, M., Abdollahpouri, H., Pechenizkiy, M., Mobasher, B., & Burke, R., 2020). La amplificación, explicada vulgarmente, se podría visualizar como el crecimiento progresivo de una bola de nieve descendiendo por la ladera de una montaña. Un pequeño trozo de hielo puede acabar siendo una enorme masa de nieve a medida que avanza por la pendiente. De la misma manera, el sesgo de un grupo reducido de la base de datos daría lugar a errores garrafales en el estudio de los datos del sistema de recomendación, como resultado de la continua repetición.

Por lo tanto, la falta de uniformidad en las elecciones de las valoraciones de los usuarios y la afectación de los sesgos comentados sobre los datos pone en entredicho la calidad de los datos a la hora de estudiar si la muestra es representativa del conjunto completo de la población o no. En el momento en el que un sistema de recomendación

basa dichas recomendaciones en datos afectados por un sesgo, y se utiliza de manera repetitiva, la amplificación del sesgo puede llevar a resultados carentes de valor y contenido. Decimos esto porque las recomendaciones propuestas por el sistema pasan a ser resultados carentes de valor y de contenido, resultados con falta de granularidad y homogéneos, que no nos aportan gran cosa a la hora de llevar a cabo un análisis sobre ellos. Los problemas que se encuentran como consecuencia de esta amplificación son, entre otros, la homogeneización de los grupos de usuarios, la falta de diversidad en cuanto a respuestas y, sobre todo, la tergiversación de la realidad frente a los datos obtenidos.

Los sistemas de recomendación son muy dinámicos y se actualizan de manera constante para reflejar las preferencias actualizadas de los usuarios en vez de una versión obsoleta. Así, supone crucial entender que la amplificación del sesgo no solo se produce, sino que se va agravando a medida que se sigue usando el modelo, dado al carácter repetitivo de éste, como comentamos anteriormente con el ejemplo de la bola de nieve.

Para abordar este problema, es fundamental que las personas que se encargan de llevar a cabo el desarrollo e implementación de los sistemas de recomendación adopten medidas para reducir el riesgo en un momento previo a que se produzca esta amplificación. Si el sesgo se evita desde un primer momento evitaríamos la reproducción del problema y los efectos serían de menor gravedad.

En conclusión, la amplificación del sesgo en los sistemas de recomendación tiene un impacto extremadamente negativo en la calidad de los resultados y puede llevar a una tergiversación de la realidad frente a los datos obtenidos.

III. ALGORITMO DE RECOMENDACIÓN DE AMAZON

a. Historia e innovación

Jeff Bezos fundó el comercio electrónico Amazon en Seattle (Washington) en 1994. Bezos, directivo de Wall Street, tomó la decisión de dejar su puesto en Nueva York y lanzar la plataforma que facilita la compraventa de productos que hoy es de sobra conocida. Con el potencial y la magnitud que el empresario preveía que tendría la firma, eligió el nombre "Amazon" como si se tratase de una imagen del río Amazonas, el más caudaloso del mundo. Pero, ¿qué es lo que diferencia a esta plataforma de otras de su sector?

Al principio, Amazon era una tienda en línea focalizada sobre todo en la venta de libros (Stone, B., 2013), producto perfecto para vender en línea por dos razones: el poco

espacio de almacenamiento que ocupa y la simple logística que supone el traslado del producto del vendedor al comprador. Como resultado, en 1995 se creó una página web de fácil manejo y acceso para cualquiera y que ofrecía una selección de libros mucho mayor que la de cualquier tienda física. Además, el portal empezó a permitir a los usuarios hacer cosas hasta entonces inauditas, como enviar reseñas de los libros que habían comprado y consultar reseñas de otros usuarios para ayudarles a decidir qué libros comprar. En consecuencia, la plataforma empezó a ganar popularidad y a expandirse, hasta que en 1997 la empresa ya había ampliado su línea de productos para incluir además de libros, herramientas y artículos para el hogar. Más adelante, Amazon empezó a introducir iniciativas innovadoras en sector tecnológico tanto en lo relativo a productos como a servicios. Por un lado, desarrolló la línea de productos electrónicos entre los que destacaron considerablemente los libros electrónicos Kindle o los altavoces inteligentes Echo. Por otro lado, lanzó su servicio de transmisión de video Amazon Prime Video (Documental de Amazon, s.f.) y su servicio de música en línea Amazon Music.

Aun así, el método innovador, original y distintivo que esta empresa de comercio electrónico ha empleado para vender sus productos y servicios debería ser el centro de nuestra atención, más que los productos y servicios en sí. El sistema de recomendación, que Amazon estrenó en 1999 y se basa en el comportamiento de los usuarios, alteró radicalmente la forma de vender. Este sistema hacía recomendaciones de bienes comparables a los que un consumidor había comprado o buscado anteriormente, lo que aumentó considerablemente las ventas y mejoró la experiencia del cliente.

Además de la forma en la que vende, Amazon destaca por su logística. La empresa ha invertido en el desarrollo tecnológico de esta importante rama de su negocio y ha conseguido mejorar la eficiencia de sus almacenes mediante robots que trabajan en sus centros o tecnología de reconocimiento de voz. Gracias a la decisión de Bezos de invertir en innovación, la red de entregas de la empresa utiliza ahora hasta drones y entregas en el mismo día. Esto, a su vez, aumenta la velocidad y la eficacia de las entregas de paquetes y, en consecuencia, la satisfacción de los clientes. Por último, pero no menos importante, cabe mencionar que Amazon lanzó su servicio de almacenamiento en la nube, Amazon Web Services ("AWS"), en 2002. Desde entonces, AWS ha crecido hasta convertirse en uno de los principales proveedores de servicios en la nube, aportando a Amazon ingresos anuales de miles de millones de dólares.

Sin embargo, de todas las iniciativas e instrumentos de innovación destacados, su sistema de recomendación es el que más impacto ha tenido en el sector. El algoritmo de

Amazon hace especial hincapié en la personalización, utilizando la IA y el aprendizaje automático para personalizar la experiencia del usuario en el sitio web y las aplicaciones móviles de la empresa y ofrecerle una sugerencia en función de su perfil.

En conclusión, al invertir en la innovación como parte integral de su empresa, Amazon ha redefinido la forma de vender.

b. Descripción del algoritmo de recomendación de Amazon

El algoritmo de recomendación de Amazon es una herramienta de aprendizaje automático que se utiliza para analizar los patrones de compra y navegación de los clientes en el sitio web de Amazon y proporcionar recomendaciones personalizadas de productos. Como hemos dicho en apartados anteriores, el objetivo de todo algoritmo de recomendación es encontrar las ponderaciones óptimas para el modelo, lo que refleja la probabilidad de que los patrones aprendidos por el modelo sean una representación precisa de las relaciones en los datos. Este tipo de algoritmo cuenta con una función de pérdida que mide las consecuencias negativas incurridas cuando las predicciones propuestas por el modelo no coinciden con los valores reales. En conclusión, el objetivo consiste en tratar de optimizar el proceso para mitigar la discrepancia entre ambos valores.

En Amazon Machine Learning, se utilizan tres funciones de pérdida diferentes, cada una para un tipo de problema de predicción respectivamente. La técnica de optimización utilizada es el Stochastic Gradient Descent (SGD) (Amazon Web Services, s.f.), que actualiza las ponderaciones de las características para cada ejemplo de forma secuencial durante cada paso a través de los datos de entrenamiento, con el objetivo de encontrar las ponderaciones óptimas que minimicen la pérdida. En términos generales, el algoritmo funciona (Franklin, L., 2019) de la siguiente manera:

En primer lugar, se lleva a cabo la recopilación de datos. En esta parte del proceso, Amazon recopila información relativa a los clientes a través de su comportamiento en su plataforma: las navegaciones, búsquedas, compras y devoluciones. Esta información no solo incluye datos acerca de transacciones, sino además datos demográficos, de geolocalización, opiniones o valoraciones de productos entre tantos otros datos que se pueden recopilar a través del uso de cookies y tecnologías similares. Esta primera parte del procedimiento tiene que ir muy en línea con la rama legal de la actividad. Esto quiere decir que solamente se podrán recopilar los datos siguiendo la normativa de protección de datos y fundamentando la recopilación de datos personales en la base de legitimación

correspondiente. Así, habrá que entender antes de recopilar los datos, qué datos se quieren recopilar y qué base de legitimación se necesita para que se de esta recopilación de datos. En línea con este último apunte, resulta de vital importancia elegir cuidadosamente el tipo de cookies que han de ser usadas a la hora de conseguir unos u otros tipos de datos. Existe un amplio rango de cookies, aunque las más habituales son las técnicas, para las que no es necesario pedir el consentimiento del interesado. Para el resto de tipos de cookies, el responsable del tratamiento de los datos, en nuestro caso Amazon, tendrá que recabar el consentimiento del individuo cuyos datos se van a recopilar y tratar.

En segundo lugar, los datos recopilados son procesados, es decir, se lleva a cabo el tratamiento y procesamiento de estos datos. Para llevar a cabo la segunda parte del procedimiento, se utilizan técnicas de aprendizaje automático y análisis de datos. Los datos recopilados no siempre son evidentes. El algoritmo puede inferir información del comportamiento que aporte información no aportada por el usuario de manera explícita. Por ejemplo, aunque no se especifique la edad, se puede entender la edad de un usuario por el tipo de productos que compra o las búsquedas que hace. De esta forma, el procesamiento incluye la creación de perfiles de clientes basados en sus patrones de compra y navegación, la identificación de patrones de compra comunes y la extracción de características relevantes de los productos. En esta parte de nuestro procedimiento, el algoritmo estudia los datos que ha recopilado, las interacciones, el comportamiento y aprende de todo ello para así sacar conclusiones predictivas en futuras interacciones o búsquedas. Esta es la etapa en la que más claramente se ve la iteración de la que hacíamos mención cuando hablamos de la amplificación del sesgo, en capítulos anteriores, dado que no es posible que se de el aprendizaje sin que el algoritmo repita una y otra vez la información que sabe con certeza. Después de haber hecho una repetición constante, se hace posible la predicción de cualquier posible situación que surja de una situación de hecho predispuesta, porque ya se han estudiado y entendido las ramificaciones probables.

En tercer lugar, una vez que los datos han sido procesados y tratados, se lleva a cabo el análisis de los mismos para poder identificar patrones y tendencias en el comportamiento de los clientes. El análisis puede incluir técnicas como la agrupación de clientes en segmentos basados en sus características, la identificación de productos que se compran con frecuencia juntos y el estudio de las opiniones de los clientes para determinar qué productos son populares. En esta parte del procedimiento, el algoritmo estudia el contexto en el que el usuario interactúa. Con esto nos referimos a que se trata de relacionar todo lo que ha aprendido en etapas anteriores con el comportamiento real

del usuario. En este sentido, por ejemplo, si el algoritmo de recomendación ha recopilado y procesado datos sobre dos usuarios distintos, que han comprado productos con el logo del Real Madrid desde que se dieron de alta en la plataforma hasta ahora, podrá llegar a la conclusión de que son ambos fervientes seguidores del equipo y los podrá clasificar en el grupo de “*madridistas*”.

En cuarto lugar, el algoritmo genera las recomendaciones para el interesado. Una vez que la información ya ha sido recopilada, procesada y analizada, se generan recomendaciones basadas en el comportamiento de compra y navegación del usuario, así como en las características de los productos, los grupos en los que se ha clasificado al usuario y las valoraciones de los clientes.

En cuarto y último lugar, se realiza la presentación de las recomendaciones pertinentes. El cliente es presentado con una serie de recomendaciones de distintas maneras. Las recomendaciones se pueden dar como una lista de productos relacionados con una búsqueda, propuestas parecidas en la misma página del producto, recomendaciones personalizadas por correo electrónico o anuncios de productos en otras páginas web. Una vez más, esta etapa necesariamente tiene que apoyarse en el aspecto legal porque el sistema solo podrá influir en el usuario lo que éste le permita. Teniendo en cuenta los ejemplos propuestos, el que podría dar mayor número de problemas son las recomendaciones personalizadas mandadas por correo electrónico al usuario en cuestión. Ese envío supone una incisión en la intimidad del individuo así que se necesitará una base de legitimación reforzada para que sea lícita. Para estar facultado a llevar a cabo este tratamiento, el responsable del mismo tiene que poder demostrar una de las dos siguientes situaciones: (i) el usuario ha dado consentimiento expreso para permitir ese envío; (ii) la empresa, Amazon, tiene un interés legítimo que resulta pesar más que el interés del usuario, después de llevar a cabo un test de ponderación de intereses.

Las etapas descritas tienen lugar gracias al papel fundamental que juega el algoritmo A9 (Hanna, R., 2018), un sistema de clasificación y organización de productos que se usa para determinar qué productos aparecen en los resultados de búsqueda y en qué orden. Este algoritmo tiene en cuenta una serie de factores para clasificar los productos de manera efectiva y proporcionar a los usuarios los productos más relevantes para su búsqueda. Entre los factores que el algoritmo A9 considera se encuentran el historial de compras del usuario, la relevancia de los términos de búsqueda utilizados por el usuario, la disponibilidad del producto, la calidad de las imágenes y descripciones del producto, el precio, las valoraciones y reseñas de los clientes, entre otros. La característica

que define al algoritmo A9 de Amazon es que se actualiza regularmente con el objetivo de mejorar la experiencia del usuario y garantizar así que los productos más relevantes y útiles aparezcan en los primeros resultados de búsqueda de cada usuario. Además de esta actualización continua, Amazon utiliza datos de ventas y compras de otros usuarios para basarse en ellos y mejorar la precisión del algoritmo. Así, el algoritmo es capaz de brindar una experiencia personalizada a cada usuario.

El acceso a las primeras posiciones en las listas de resultados es altamente valioso para las marcas, ya que se estima que en 2019 el 35% de los usuarios hizo clic en el primer producto que aparecía, y el 66% de los usuarios hizo clic en uno de los primeros productos de la primera página de resultados (BBVA, s.f.). Por lo tanto, alcanzar los primeros puestos en las listas de resultados debe ser un objetivo prioritario para las marcas que buscan vender en Amazon. Para lograrlo, se deben tener en cuenta varias variables que, cuando se aplican de manera conjunta, pueden ayudar considerablemente a conseguir el éxito en ventas. El vendedor tendrá que prestar especial atención a factores (Bowman, B., s.f.) como el título, la foto o la descripción de producto a la hora de presentarlo en la plataforma, de forma que se pueda maximizar la posibilidad de beneficios.

IV. ANÁLISIS DE LOS DATOS

Hemos hablado durante el desarrollo de este trabajo cómo los algoritmos que utilizan las empresas para hacer un filtrado de los clientes para tratar de ofrecer recomendaciones personalizadas de productos descubriendo patrones en las interacciones entre consumidores y productos pueden resultar contraproducentes en caso de que exista un sesgo en la base de datos. Este sesgo se ha estudiado de forma teórica, pero en este capítulo desarrollaremos de forma práctica los efectos del sesgo sobre los sistemas de recomendación.

Para ello, utilizaremos una base de datos que recoge información sobre los comportamientos de compraventa de distintos usuarios en la plataforma de Amazon. La base de datos en concreto contiene información sobre más de 1.000.000 de individuos, dividida en 10 campos distintos.

Las primeras dos variables (“item_id” y “user_id”) contienen el número que identifica a los distintos productos que han sido estudiados, así como el identificador de cada usuario. La tercera y cuarta, (“rating” y “timestamp”) recogen las distintas valoraciones que ha hecho cada usuario para el producto determinado, y la fecha específica de cada compraventa. La quinta (“model_attr”) es de especial importancia para

nuestro estudio puesto que refleja el adjetivo que se le ha atribuido a cada producto a la hora de presentarlo en la plataforma. La sexta y séptima (“category” y “brand”) se refieren al producto y establecen la categoría en la que se encasilla cada uno y la marca. Las últimas dos variables (“year” y “user_attr”) enseñan el año en el que se llevó a cabo la transacción y el atributo real del usuario que compró el producto, que tendremos que tener en cuenta para nuestro estudio.

Esta base de datos nos puede ser tremendamente útil a la hora de desarrollar el tema por diferenciarse un evidente sesgo en cuanto al género. Como hemos mencionado en el párrafo anterior, la base de datos clasifica los productos en los dos géneros a los que va dirigido cada producto (mujer y hombre), y se comercializan acorde a dicha categorización. De esta manera, entendemos a través de la base de datos elegida que las estrategias de marketing que ponen el foco en específicos grupos de personas producen un set de datos que, por resultar sesgado, supone un peor sistema de recomendación por parte de empresas como Amazon. Esto supone una potencial pérdida para los vendedores, una experiencia insatisfactoria por parte del usuario y, por último, pero no menos importante, da lugar a un preocupante problema social.

A la hora de importar los datos a R, había un problema con alguna de las variables de la base de datos. Inicialmente la base contaba con exactamente 1.048.575 de observaciones, sin embargo, decidimos extraer una muestra más manejable para el estudio de los datos mediante la supresión de algunas de las observaciones. Para tomar la decisión de cuáles de las filas serían eliminadas, nos fijamos en la variable de “user_attr”, que refleja el atributo real del usuario que ha comprado el producto en cuestión, y nos dimos cuenta de que había un gran número de campos vacíos en ella. Teniendo esto en cuenta utilizamos la siguiente fórmula para extraer una muestra de la base de datos en la que solamente se incluyeran las observaciones de las cuales se supiera el atributo real del usuario. Los pasos a seguir fueron los siguientes:

En primer lugar, importamos la base de datos completa con todos los datos y la llamamos “tabla”.

FIGURA 1: Tabla descriptiva de la base de datos

	item_id	user_id	rating	timestamp	model_attr	category	brand	year	user_attr
1	0	0	5.0	1999-06-13	Female	Portable Audio & Video	NA	1999	NA
2	0	1	5.0	1999-06-14	Female	Portable Audio & Video	NA	1999	NA
3	0	2	3.0	1999-06-17	Female	Portable Audio & Video	NA	1999	NA
4	0	3	1.0	1999-07-01	Female	Portable Audio & Video	NA	1999	NA
5	0	4	2.0	1999-07-06	Female	Portable Audio & Video	NA	1999	NA
6	0	5	2.0	1999-07-12	Female	Portable Audio & Video	NA	1999	NA
7	0	6	5.0	1999-07-13	Female	Portable Audio & Video	NA	1999	NA
8	0	7	2.0	1999-07-13	Female	Portable Audio & Video	NA	1999	NA
9	0	8	4.0	1999-07-16	Female	Portable Audio & Video	NA	1999	NA
10	0	9	5.0	1999-08-20	Female	Portable Audio & Video	NA	1999	NA
11	0	10	1.0	1999-08-24	Female	Portable Audio & Video	NA	1999	NA

Fuente: Elaboración propia

Después, eliminamos los valores que no nos servían, es decir, los campos vacíos dentro de la variable “user_attr” y creamos la muestra de datos a la que llamamos “ntabla”.

FIGURA 2: Tabla descriptiva de base de datos “limpia”

	item_id	user_id	rating	timestamp	model_attr	category	brand	year	user_attr
1	0	28	2.0	1999-12-01	Female	Portable Audio & Video	NA	1999	Female
2	3	81	5.0	2000-03-31	Female	Camera & Photo	NA	2005	Male
3	7	131	4.0	2000-06-15	Female	Home Audio	Philips	2001	Male
4	3	139	3.0	2000-06-29	Female	Camera & Photo	NA	2005	Female
5	8	178	1.0	2000-10-19	Female	Accessories & Supplies	NA	2000	Male
6	19	231	5.0	2000-12-31	Female	Camera & Photo	Canon	2000	Male
7	14	233	5.0	2001-01-02	Female&Male	Camera & Photo	Kodak	2001	Female
8	14	257	5.0	2001-02-04	Female&Male	Camera & Photo	Kodak	2001	Female
9	14	269	5.0	2001-03-02	Female&Male	Camera & Photo	Kodak	2001	Male
10	23	274	5.0	2001-03-10	Female&Male	Camera & Photo	NA	2001	Male
11	15	269	5.0	2001-03-10	Female&Male	Camera & Photo	Kodak	2006	Male

Fuente: Elaboración propia

De esta manera, nos aseguramos de tener una muestra de los datos en la que todas las observaciones tuvieran registradas un atributo real para sus usuarios y redujimos el número total de filas a 144.408.

a. ¿Existe el sesgo?

A la hora de detectar si, efectivamente, existe el sesgo, el paso de extraer una muestra dentro de una base de datos compleja puede ser de gran utilidad para examinar

los datos. Con una muestra de menos de 200.000 observaciones resulta más fácil revisar las observaciones y sus 10 variables y será más manejable para entender el sesgo.

i. Proceso de selección del producto

Después de inspeccionar las diez variables, hemos identificado que las dos con mayor posibilidad de estar sujetas a sesgos en esta parte del sistema de recomendación son “model_attr” y “user_attr”. La relación que exista entre la categoría en la que se ha clasificado al usuario y la categoría real de éste será de crucial importancia para evaluar si existe sesgo en la base de datos o no. Nuestra hipótesis inicial es que no existirá una diferencia sustancial entre ambas variables y, por lo tanto, se demostrará el enorme efecto que tiene la asignación de adjetivos a los productos, y sus consecuentes implicaciones, en la elección de los usuarios.

Por un lado, a la hora de entender esta relación, una buena herramienta es el test de Chi-Cuadrado entre las dos variables. El código en R quedaría de la siguiente manera:

FIGURA 3: Test de Chi-Cuadrado

```
> resultado_chi <- chisq.test(ntabla$model_attr, ntabla$user_attr)
> resultado_chi
```

Pearson's Chi-squared test

```
data: ntabla$model_attr and ntabla$user_attr
X-squared = 462.83, df = 2, p-value < 2.2e-16
```

Fuente: Elaboración propia

El resultado del análisis de Chi-Cuadrado de ese estudio es el siguiente. En este caso, se establece la hipótesis nula asumiendo que las variables "model_attr" y "user_attr" no tienen ninguna relación o asociación significativa. Por lo tanto, siguiendo la hipótesis nula, las distribuciones de categorías en ambas variables en la población objetivo son independientes una de la otra. Una vez entendido esto, pasaremos a analizar el resultado obtenido.

En primer lugar, los grados de libertad (df) son dos. Los grados de libertad son la proporción de categorías independientes en los datos menos las restricciones. Hay dos niveles de libertad en este caso.

En segundo lugar, el valor p-value es menor a $2.2e-16$. La probabilidad de obtener un valor Chi-Cuadrado igual o más extremo que el valor observado se conoce como valor p si se asume que la hipótesis nula es verdadera. Por lo tanto, se entiende que la probabilidad de obtener un valor de Chi-Cuadrado tan extremo o más extremo que el observado es extremadamente baja en este caso, ya que el valor p es menor que $2.2e-16$. Como resultado, se rechaza la hipótesis nula.

En tercer lugar, el valor de Chi-Cuadrado (X-squared) es de 462.83. En un análisis de Chi-Cuadrado, el valor de Chi-Cuadrado se utiliza para medir la discrepancia entre los datos observados y los datos esperados, bajo la hipótesis nula. De esta forma, se da una relación directamente proporcional entre la discrepancia entre los datos mencionados y el valor de Chi-Cuadrado.

Después de estudiar todas las soluciones obtenidas, se ha llegado a la conclusión de que hay suficiente evidencia para sugerir que existe una asociación significativa entre las variables, ya que el valor p es tan pequeño que se ha rechazado la hipótesis nula. En definitiva, los resultados del análisis de Chi-Cuadrado muestran que, en la muestra examinada, las variables "model_attr" y "user_attr" tienen una relación significativa, lo que reafirma nuestra idea inicial.

Por otro lado, se puede estudiar la relación entre variables mediante una tabla de contingencia de la distribución de frecuencias de el adjetivo designado y la elección de los usuarios. El resultado obtenido de RStudio es el siguiente:

FIGURA 4: Tablas de contingencia y proporción (user_attr y model_attr)

```
> tabla_contingencia <- table(ntabla$model_attr, ntabla$user_attr)
> tabla_contingencia
```

	Female	Male
Female	28567	26553
Female&Male	20742	20153
Male	21987	26406

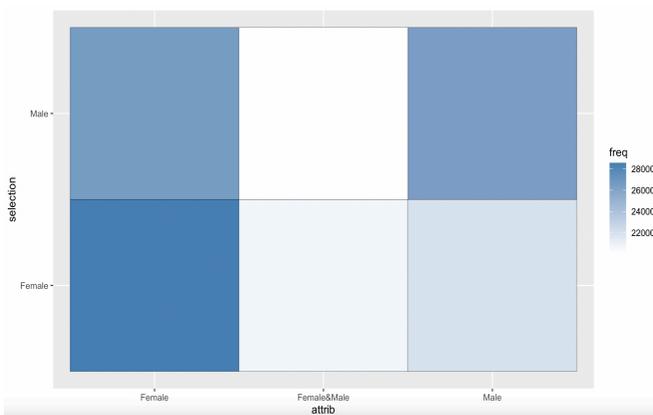
```
> frecuencias_relativas <- prop.table(tabla_contingencia)
> frecuencias_relativas
```

	Female	Male
Female	0.1978215	0.1838749
Female&Male	0.1436347	0.1395560
Male	0.1522561	0.1828569

Fuente: Elaboración propia

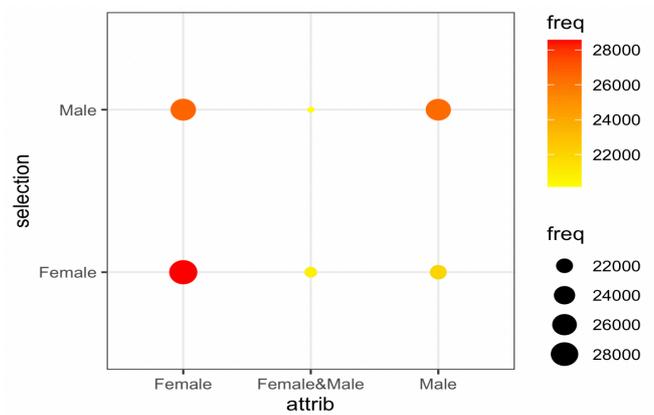
De esto podemos sacar una serie de conclusiones. Para empezar, que el total de designaciones a un género específico, hombre o mujer, es conjuntamente mucho más elevado que el número de asignaciones “unisex”, las asignaciones “*Female&Male*”. A un total del 33.5% de los productos se les ha asignado la etiqueta masculina, y a un 38.2% la etiqueta femenina, lo que suma un 71.7% de las asignaciones. Mientras tanto, solo quedan un 28.3% de productos con una etiqueta un poco menos definida. Además de esta observación, vemos que nuestra base de datos esta compuesta por un 49.4% de mujeres y un 50.6% de hombres. Un gráfico puede ser de utilidad para ayudar al lector a visualizar la tabla de contingencia presentada:

FIGURA 5: *Heatmap de la tabla de contingencia*



Fuente: *Elaboración propia*

FIGURA 6: *Gradiente de la tabla de contingencia*



Fuente: *Elaboración propia*

Gracias a estas representaciones de la relación entre la atribución de adjetivos que se le ha dado a cada producto y la selección llevada a cabo por el usuario, podemos ver la frecuencia entre ambas. Por un lado, vemos como el 19.78% de los usuarios femeninos eligen productos que van dirigidas a este segmento, porcentaje muy superior al de los usuarios femeninos que deciden adquirir productos que se podrían considerar “unisex” o productos etiquetados como masculinos. De la misma manera vemos la asociación entre la selección de usuarios masculinos de estos productos y el etiquetado masculino.

Inicialmente, se ha planteado la hipótesis de que existiría una relación muy cercana entre el adjetivo atribuido al producto en su presentación en la plataforma y el atributo real del usuario que adquiere el producto. Esta relación sería una demostración de que existe un sesgo de marketing en la base de datos. Gracias a las dos pruebas estadísticas llevadas a cabo, podemos llegar a la conclusión de que, efectivamente, estas

dos variables están estrechamente relacionadas y, por lo tanto, se demuestra el sesgo que intuíamos anteriormente.

ii. Nivel de satisfacción del producto

El nivel de satisfacción es un reflejo de lo satisfecho que se ha quedado el cliente con el producto comprado. Para estudiarlo, empezaremos haciendo una tabla de contingencia entre el atributo del comprador y los ratings aportados.

FIGURA 7: Tablas de contingencia y proporción (user_attr y rating)

```
> tabla_contingencia <- table(ntabla$user_attr, ntabla$rating)
> tabla_contingencia
```

	1.0	2.0	3.0	4.0	5.0
Female	5281	3657	5729	11734	44895
Male	4952	3933	6459	14020	43748

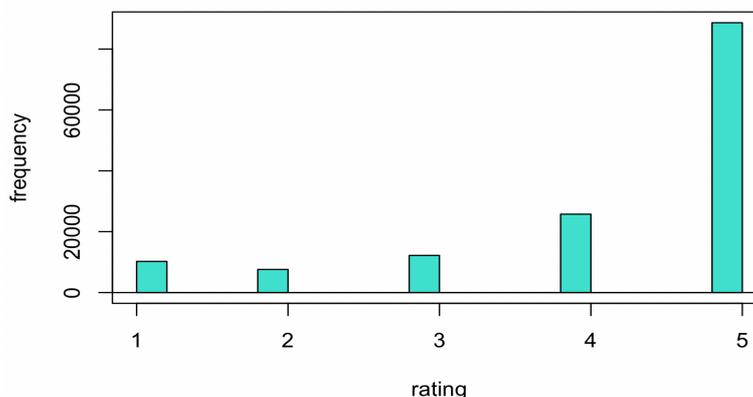
```
> frecuencias_relativas<-prop.table(tabla_contingencia)
> frecuencias_relativas
```

	1.0	2.0	3.0	4.0	5.0
Female	0.03657000	0.02532408	0.03967232	0.08125589	0.31088998
Male	0.03429173	0.02723533	0.04472744	0.09708603	0.30294721

Fuente: Elaboración propia

A la hora de visualizar estos datos, utilizamos un histograma que refleje la relación entre cada rating y el número de veces que ha sido votado semejante rating. De esta manera, podremos ver el rating más votado, en general.

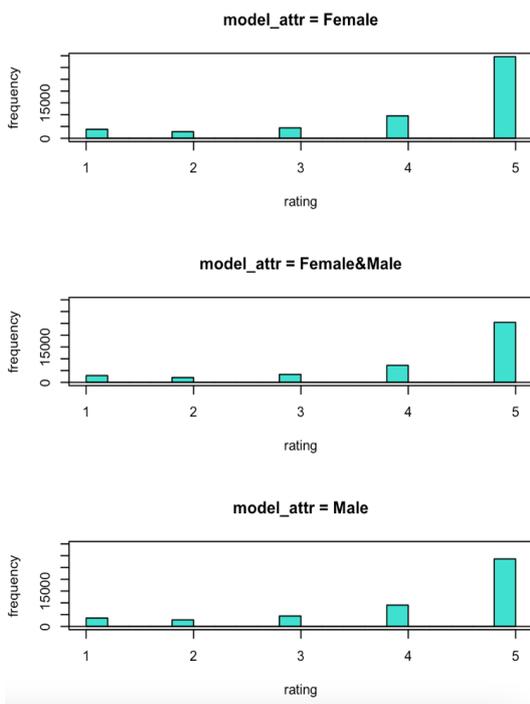
FIGURA 8: Histograma de frecuencia sobre los ratings



Fuente: Elaboración propia

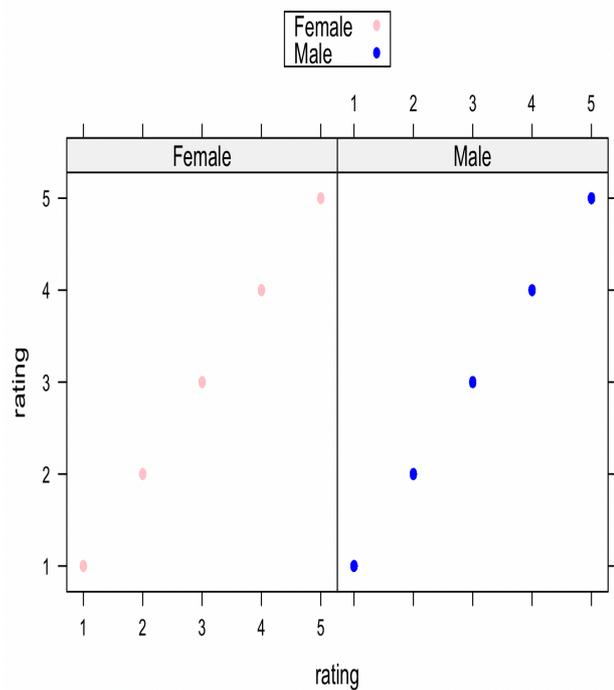
Vemos así cómo parece que el rating más votado normalmente tanto por usuarios masculinos como femeninos es el 5, lo que nos llevaría a pensar que ciertamente los usuarios están satisfechos con los productos comprados. Las hipótesis que explicaremos en este apartado se pueden resumir en dos. Se espera que exista una gran similitud entre los ratings según el género de los usuarios ya sea estudiado desde el punto de vista de los usuarios reales (user_attr) o de los designados (model_attr). Asimismo, entendemos que debe existir una cercana similitud entre los ratings de las distintas categorías de productos. De forma general, se puede ver esta similitud en el ámbito de la identidad del usuario a través de estos gráficos.

FIGURA 9: Histograma comparativo de los ratings de cada categoría de model_attr



Fuente: Elaboración propia

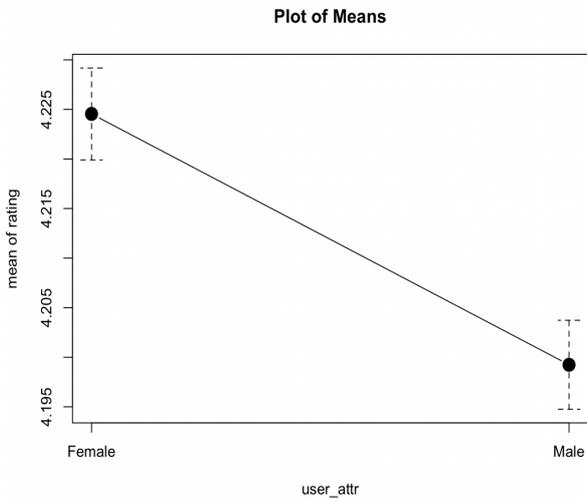
FIGURA 10: XYPlot comparativo de los ratings de cada categoría de user_attr



Fuente: Elaboración propia

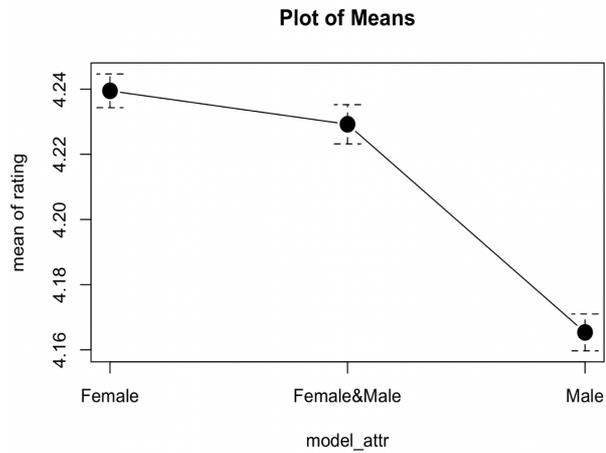
Aun así, si estudiamos la diferencia entre la media de las valoraciones de ambos grupos, se puede ver de manera más exacta:

FIGURA 11: Comparativa gráfica de la media en los ratings de cada categoría de user_attr



Fuente: Elaboración propia

FIGURA 12: Comparativa gráfica de la media en los ratings de cada categoría de model_attr

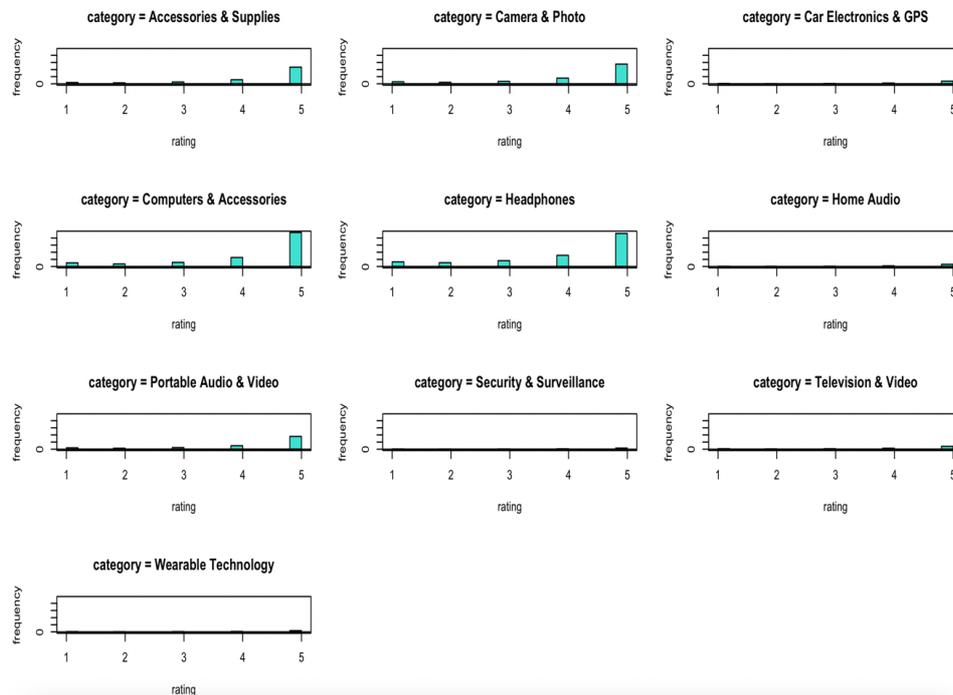


Fuente: Elaboración propia

Aunque en los dos primeros gráficos se podía apreciar bastante bien la semejanza entre las valoraciones de los grupos, gracias a los segundos vemos como la diferencia entre la media de las valoraciones de los distintos sectores es casi ínfima, siendo ésta de apenas centésimas.

De todas maneras, hay que llevar el análisis un poco más allá y estudiar los ratings según la categoría para entender un poco mejor. Para ello llevamos a cabo una representación de los ratings filtrada según la categoría del producto y conseguimos el siguiente resultado.

FIGURA 13: Comparativa gráfica de los ratings de cada categoría de producto



Fuente: Elaboración propia

Este análisis, además de visualmente, se puede demostrar de manera numérica a través del estudio de la varianza (ANOVA). En un análisis de este estilo se cuenta con dos variables, una que es la que nos interesa estudiar, la llamaremos variable dependiente y otra que es independiente, la llamaremos variable independiente (Madrigal, L., 2012). Lo normal es que la variable dependiente sea una numérica continua, lo que posibilita al investigador a obtener una gama muy amplia de resultados. En nuestro caso, la variable dependiente es el valor que se le da a las valoraciones, es decir, cualquier valor entre 1 y 5, mientras que la variable independiente será el grupo por el que estemos filtrando los datos: la identidad real del usuario (`user_attr`) o la identidad designada (`model_attr`). Una vez aplicado este modelo en R, los resultados obtenidos deberían de reflejar de manera exacta lo que ya hemos visualizado a través de los gráficos, una ínfima diferencia entre las medias de las valoraciones de los distintos grupos.

FIGURA 14: Comparativa numérica de la media en los ratings de cada categoría de user_attr

	mean	sd	data:n
Female	4.224543	1.237527	71296
Male	4.199242	1.213177	73112

Fuente: Elaboración propia

FIGURA 15: Comparativa numérica de la media en los ratings de cada categoría de model_attr

	mean	sd	data:n
Female	4.239478	1.214035	55120
Female&Male	4.229221	1.217679	40895
Male	4.165354	1.243107	48393

Fuente: Elaboración propia

FIGURA 16: Comparativa numérica de la media en los ratings de cada producto

	mean	sd	data:n
Accessories & Supplies	4.361948	1.114612	17450
Camera & Photo	4.261329	1.187233	22045
Car Electronics & GPS	4.228734	1.244993	2986
Computers & Accessories	4.261536	1.202995	37406
Headphones	4.09258	1.280687	41102
Home Audio	4.24116	1.184226	2475
Portable Audio & Video	4.25587	1.203618	14169
Security & Surveillance	4.052391	1.322151	1418
Television & Video	4.087866	1.312857	3579
Wearable Technology	3.805962	1.444540	1778

Fuente: Elaboración propia

Efectivamente, vemos cómo la diferencia entre las medias es mínima. Después de llevar a cabo nuestro análisis podemos decir, para empezar, que la relación entre el nivel de satisfacción del cliente y los factores de marketing es, desde luego, complicada. De todas maneras, queda claro después de estudiar la base de datos que los usuarios quedan satisfechos con sus compras, pero esto no nos hace rechazar la hipótesis inicial de sesgo. Aunque la compra les haya resultado útil, la base de datos puede estar sesgada de todas maneras.

Como hemos mencionado ya, los sistemas de recomendación se basan sobre todo en datos históricos disponibles de los usuarios y en el perfilado llevado a cabo a raíz de su interacción con los productos. Por lo tanto, si anteriormente ya se había producido algún tipo de sesgo en la selección del producto, las recomendaciones que se le habrán hecho al usuario lo único que hacen es perpetuar semejante sesgo. Esto no quiere decir que el usuario quede insatisfecho con un producto coherente y relacionado con sus compras anteriores, lo que quiere decir es que la base de datos, con un sesgo de marketing importante, seguirá perpetuando dicho sesgo con cada compra de productos similares.

En línea con lo anterior, el sesgo en un sistema de recomendación tiene un efecto amplificador que hace que se confirmen las preferencias y comportamientos previos de los usuarios. En consecuencia, sus preferencias históricas, influenciadas por imágenes de productos cuyas campañas de marketing se han focalizado en géneros específicos, se ven reforzadas. Por lo tanto, es lógico que los usuarios otorguen ratings de valores altos por estar estrechamente relacionadas las recomendaciones con sus preferencias, pero el sesgo de marketing inicial no deja de existir.

Por último, podemos explicar el nivel de satisfacción obtenido en el sistema de recomendación por la enorme influencia que ejerce el contexto sobre el usuario. Hemos visto este tipo de sesgo anteriormente y lo que queremos decir con esto es que los usuarios pueden sentirse satisfechos con la recomendación ofrecida, pero esta satisfacción está distorsionada. Distorsionada porque las recomendaciones se ajustan a lo que esperan encontrar, aunque esto suponga un sesgo de género. Esto puede dificultar la detección del sesgo por parte de los usuarios y perpetuar el ciclo de recomendaciones sesgadas.

V. MITIGACIÓN DEL SESGO

Como hemos podido deducir de los capítulos anteriores, los sistemas de recomendación no son inmunes a los sesgos, que pueden tener consecuencias imprevistas, como prácticas discriminatorias o predicciones inexactas. Por lo tanto, es crucial mitigar los sesgos en la recomendación para garantizar que estos sistemas sean justos. Para lograr este objetivo, se ha considerado (Shekhar, G., 2021) que existen varios retos que deben abordarse con carácter previo al uso de una base de datos para un sistema de recomendación.

El primer reto está relacionado con la recopilación de datos. En muchos casos, puede que no haya suficientes datos disponibles para los segmentos que están infrarrepresentados, lo que puede dar lugar a ese sesgo de selección del que hablábamos antes. Por ejemplo, si un sistema de recomendación se entrena con datos que representan mayoritariamente a un sexo, puede que no funcione bien con datos del otro sexo. Además, el uso de datos de origen colectivo se ha hecho más popular, y es importante tener en cuenta los sesgos individuales de los creadores de datos. Por ejemplo, un conjunto de datos puede llegar a reflejar los prejuicios de las personas que enviaron los datos, lo que daría lugar a un modelo injusto.

El segundo reto consiste en identificar las características sensibles que pueden causar sesgos involuntarios (Deldjoo, Y., Bellogin, A., & Di Noia, T., 2021). Las

características sensibles son atributos altamente correlacionados con determinadas características protegidas, como la raza o el sexo. Por ejemplo, si un sistema de recomendación utiliza códigos postales para predecir la aprobación de préstamos, puede discriminar inadvertidamente a determinados grupos raciales que tienen más probabilidades de vivir en códigos postales específicos. En la actualidad, no existe un método estándar para detectar estas características y, por lo tanto, debemos basarnos en el conocimiento del dominio y en el análisis exploratorio.

El tercer reto se refiere al creciente uso de datos multimodales (Peska, L., 2017), que incluyen texto, imagen y vídeo. Los enfoques existentes para detectar y mitigar los sesgos están diseñados principalmente para datos estructurados, lo que significa que las capacidades de detección de sesgos deben reforzarse para manejar datos multimodales. Por ejemplo, un sistema de recomendación que analiza imágenes médicas para diagnosticar una enfermedad también puede tener en cuenta los historiales de los pacientes basados en texto. Sin embargo, si el sistema está sesgado en contra de ciertos grupos de pacientes, puede diagnosticar o tratar mal a esos pacientes.

El último reto está relacionado con la *feature engineering* (Garg, N., 2022), que mejora el rendimiento del modelo. La *feature engineering* consiste en seleccionar y transformar los datos de entrada para mejorar la precisión del modelo. Sin embargo, las características complejas pueden ser difíciles de rastrear. Este problema puede dar lugar a modelos difíciles de interpretar, lo que dificulta la identificación y corrección de sesgos.

En conclusión, mitigar los sesgos en la recomendación es esencial para crear sistemas fiables y justos (Berkovsky, S., Kuflik, T., & Ricci, F., 2011). Abordar los retos relacionados con la recopilación de datos, la identificación de características sensibles, el manejo de datos multimodales y la ingeniería de características es fundamental para lograr una base de datos de calidad.

En línea con esto último, una vez tenemos una base de datos, existen diferentes vías para verificar esta calidad en la base de datos con carácter posterior al estudio de los retos que acabamos de describir. Para ello, diferenciaremos una serie de indicadores que nos ayuden en la tarea.

VI. CALIDAD DE LA BASE DE DATOS

a. Concepto

Evaluar la calidad de la base de datos de un sistema de recomendación es un paso crucial para mejorar su eficacia y aumentar la satisfacción del usuario. Además, para evitar sesgos en la calidad de los datos, es necesario garantizar que los datos utilizados en los sistemas de recomendación sean fiables, precisos y representativos de los usuarios. Las dos dimensiones principales para medir la calidad de una recomendación son la precisión y la novedad y diversidad (Cañamares Pérez, R., 2019).

Las medidas de precisión se utilizan para evaluar el nivel de éxito de una recomendación basándose en la opinión del usuario sobre los elementos recomendados. Existen dos tipos de medidas de precisión: medidas de error y medidas de clasificación. Las métricas de error han sido las más utilizadas durante mucho tiempo y se centran en el valor de valoración predicho por el algoritmo y en su proximidad a la valoración real asignada por el usuario al elemento objetivo. Algunos ejemplos de medidas de error son el error medio absoluto (MAE) y el error cuadrático medio (RMSE). Sin embargo, estas medidas no siempre son representativas de la satisfacción real del usuario y han sido sustituidas por medidas que se correlacionan mejor con la utilidad final para los consumidores reales.

Las medidas de novedad y diversidad se utilizan para evaluar la recomendación de artículos que los usuarios pueden no conocer y para aumentar las recomendaciones de artículos disímiles. Las medidas de novedad recompensan la recomendación de artículos que no se recomiendan habitualmente, mientras que las medidas de diversidad se centran en la recomendación de artículos disímiles. Algunos ejemplos de medidas de novedad y diversidad son la novedad de cola larga, lo inesperado, la diversidad dentro de la lista y la diversidad de ventas.

Además de las medidas de precisión, novedad y diversidad, se utilizan otras dimensiones para medir el rendimiento de las empresas, como el número de compras, el tamaño de la cesta y el porcentaje de clics. Estas medidas complementan las de precisión, novedad y diversidad para medir la eficacia global del sistema de recomendación y algunos ejemplos son (Adomavicius, G., Bockstedt, J. C., Curley, S. P., & Zhang, J., 2019) la limpieza de datos, la imputación de datos, el aumento de datos y el muestreo de datos. La limpieza de datos implica la eliminación o corrección de datos erróneos, incoherentes o duplicados, como valores atípicos, errores tipográficos o anomalías. La

imputación de datos consiste en completar los valores que faltan con estimaciones adecuadas, como medias, medianas o predicciones. El aumento de datos consiste en complementar los datos con información adicional, como metadatos, características o etiquetas. El muestreo de datos consiste en seleccionar un subconjunto de datos que refleje la diversidad y distribución de los usuarios y elementos, como el muestreo estratificado, el muestreo aleatorio o el muestreo por conglomerados.

En general, la evaluación de los sistemas de recomendación resulta esencial para mejorar su calidad y aumentar la satisfacción de los usuarios. A la hora de llevar a cabo semejante evaluación las principales medidas que se deben utilizar son las de precisión y novedad. Aun así, otras medidas como las de rendimiento comercial pueden ser de gran utilidad por la información complementaria que aportan para analizar la eficacia del sistema de recomendación.

b. Soluciones

Hemos visto la importancia de la calidad de los datos en los sistemas de recomendación desde un punto de vista más teórico, pero resulta interesante también ver el efecto que tiene sobre el rendimiento de los sistemas y el nivel de satisfacción del consumidor (Sar Shalom, O., Berkovsky, S., Ronen, R., Ziklik, E., & Amihod, A., 2015). Los principales problemas que contribuyen a una pobre calidad de los datos son la dispersión y redundancia de datos. Se han desarrollado diferentes modelos para intentar determinar soluciones factibles a estos problemas. En el ensayo mencionado se proponen modelos para establecer umbrales dependientes de los datos y distintos tipos de muestreo y ponen estas soluciones en práctica para analizar diferentes conjuntos de datos. Pero, como los autores, para explicar mejor las soluciones las desarrollaremos según sus respectivos problemas.

Por un lado, el modelo propuesto para evitar la dispersión de datos en un sistema de recomendación es el modelo de umbral. Este modelo tiene en cuenta la gran cantidad de usuarios de un sistema de recomendación que no aportan valoraciones o ratings después de llevar a cabo su adquisición. Así, la solución propuesta consiste en llevar a cabo una limpieza por producto de las valoraciones y extraer la media de los ratings existentes de un específico producto para construir un modelo efectivo. El desafío que presenta esta solución consiste en saber establecer un criterio efectivo para hacer esta limpieza y quedarse con ratings que aporten valor al modelo.

Por otro lado, se plantea el problema que supone la redundancia de los datos en las bases de datos de los sistemas de recomendación. Para paliar estos problemas los autores sugieren elegir la muestra aleatoria de los datos con menor tasa de muestreo, para así conseguir un modelo que se acerque lo máximo posible a uno que hubiera sido más rico en cuanto a variedad de datos. En línea con el desafío que surgía en el planteamiento de la solución anterior, el reto que entendemos que puede surgir de un muestreo aleatorio está en saber predecir la tasa de muestreo.

Además de estos autores, son muchos los que han propuesto sus propias soluciones para conseguir una base de datos de calidad. Así, en el estudio llevado a cabo de manera conjunta por Leo L. Pipino, Yang W. Lee, Richard Y. Wang (2002), define como las más útiles a las siguientes funciones: la de ratio simple, la de la operación de mínimo o máximo y la de la media ponderada. La tarea más complicada para estos autores es la definición de una dimensión de manera que su aplicación encaje con el objetivo empresarial que se persigue.

VII. CONCLUSIÓN

Los algoritmos que sustentan los sistemas de recomendación llevan a cabo un filtrado que permite perfilar a los diferentes tipos de usuarios existentes. De entre todos los tipos de filtrado, en este trabajo nos hemos centrado en el de *collaborative filtering*, una técnica que centra su análisis en las características no solo de los usuarios objetivos sino, además, de los terceros usuarios de la plataforma. Este método de perfilado es singular para las características implícitas que se pueden inferir de la comparativa entre el comportamiento del usuario objetivo y el de los terceros.

Esta personalización de las ventas tiene sin duda enormes ventajas como el aumento de beneficios para el vendedor o un nivel de satisfacción mayor para los usuarios que utilizan la plataforma.

Pero, el hecho de que el sistema de recomendación se base en una reiterativa comparación de interacciones, aunque denote automatización del modelo, no hace que el algoritmo sea completamente imparcial. Esto es porque la fuente de los datos utilizados es humana y, por lo tanto, inexacta y sesgada.

A lo largo de la investigación, se ha llevado a cabo una extensa exposición del concepto de sesgo y todas las distintas ramificaciones que de este concepto se desarrollan. Entre las variadas y abundantes definiciones y categorizaciones encontradas, hemos podido clasificar los diferentes tipos de sesgo en sesgos de exposición, conformidad,

selección, posición y popularidad. Aun así, la clasificación no ha sido lo más intrigante del estudio del sesgo en los sistemas de recomendación. Sin duda lo más reseñable de cualquier tipo de sesgo es el impacto y la consecuente amplificación que supone en el funcionamiento del sistema de recomendación.

Una vez descrita la parte teórica de la investigación, escogimos una base de datos ad hoc para demostrar el sesgo en los sistemas de recomendación. La base de datos contenía información sobre interacciones de usuarios con productos ofrecidos en la plataforma de Amazon. El foco de las observaciones de la base de datos estaba centrado en un tipo específico de usuario, los designados masculinos o femeninos. Después de realizar un análisis sobre los datos en relación al proceso de selección y el nivel de satisfacción del usuario con el producto, llegamos a la conclusión de que, efectivamente, la designación de una clase específica de usuario para cada uno de los productos producía un impacto en las recomendaciones posteriores.

Aun así, existen numerosas maneras de mitigar este sesgo y han sido estudiadas en el capítulo quinto de este trabajo. La clave, según lo que hemos podido entender de la literatura académica utilizada está en cuidar una serie de desafíos. Los cuatro desafíos principales son los relacionados con la selección de los datos, la identificación de características sensibles, el uso de datos multimodales y la *feature engineering*. Todas estas medidas preventivas resultan esenciales para crear una base de datos de calidad, que enriquezca y mejore el modelo del sistema de recomendación.

Este trabajo ha descrito el efecto de las campañas de marketing sesgadas en algoritmos de collaborative filtering usados en sistemas de recomendación. Sin embargo, como trabajo de investigación nos hemos encontrado con una serie de limitaciones que plantean la necesidad de llevar a cabo un estudio más profundizado sobre el tema.

Para empezar, los datos utilizados son sin duda relevantes para reflejar el problema planteado, pero no ofrecen una imagen del todo completa. El acceso a una base de datos de mayor envergadura puede resultar fundamental para llevar a cabo una investigación más exhaustiva y representativa. Por lo tanto, una buena manera de aportar más valor al trabajo sería mediante el acceso a bases de datos de tamaño mayor y que contengan datos de usuarios actualizados.

Además, la generalización suele empobrecer, lo que quiere decir que los hallazgos específicos relacionados con el sesgo de marketing en algoritmos de *collaborative filtering* pueden no ser del todo extrapolables a otros sistemas de recomendación.

Las limitaciones que presenta el trabajo que hoy nos ocupa no son más que oportunidades de mejora y potenciales líneas de investigación futura sobre esta materia. De esta manera, se proponen las siguientes sugerencias.

En primer lugar, resultaría de gran interés estudiar un amplio rango de algoritmos de collaborative filtering en lo que a sesgo de marketing se refiere. Así, se podría realizar un contraste entre ellos y conseguir una comprensión más amplia sobre la sensibilidad de los sistemas de recomendación, basados en estos algoritmos, al sesgo.

En segundo lugar, se podrían plantear políticas de marketing más éticas, de tal manera que se trate de minimizar el sesgo. El establecimiento de pautas específicas que trataran de luchar contra el sesgo, podría influir profundamente en la disminución del sesgo en los algoritmos de recomendación.

En tercer lugar, sería de gran utilidad llevar a cabo una investigación más exhaustiva con métodos empíricos. El estudio del efecto del sesgo en la confianza y satisfacción de los usuarios de los sistemas de recomendación a largo plazo, aportaría información de gran valor para incrementar la concienciación de las empresas sobre el tema.

En cuarto y último lugar, habría que investigar la manera en la que mejorar la transparencia de los algoritmos de los sistemas de recomendación. Al aumentar la transparencia en las plataformas, la decisión final estaría en manos de los usuarios, que comprenderían mejor los factores que influyen en las recomendaciones y podrían configurar más adecuadamente su decisión.

En definitiva, los sistemas de recomendación se pueden ver afectados por el sesgo humano en tanto en cuanto su base de datos se sustenta en el comportamiento de los usuarios. De esta manera, la evaluación de la calidad de la base de datos supone un paso crucial para incrementar la eficacia y prevenir los sesgos en los sistemas de recomendación. Aun así, todavía queda mucho campo de investigación sobre la materia y habrá que seguir indagando para tratar de poner solución a este acuciante problema.

VIII. BIBLIOGRAFÍA

- Abdollahpouri, H., Burke, R., & Mobasher, B. (2019). *Managing Popularity Bias in Recommender Systems with Personalized Re-ranking*. AAAI Florida Artificial Intelligence Research Society (FLAIRS '19), May 18–22, 2019, Sarasota, Florida, USA (pp. 1-6). ACM, New York, NY, USA.
- Abdollahpouri, H., Mansoury, M., Burke, R., Mobasher, B., & Malthouse, E. (2021). User-centered Evaluation of Popularity Bias in Recommender Systems. Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization (UMAP '21), June 21–25, 2021, Utrecht, Netherlands (pp. 1-12). ACM, New York, NY, USA.
- Adomavicius, G., Bockstedt, J. C., Curley, S. P., & Zhang, J. (2019). *Reducing Recommender Systems Biases: An Investigation of Rating Display Designs*. Indiana University Bloomington.
- Amazon Web Services. (s.f.). *Amazon Machine Learning: Guía para desarrolladores*. Recuperado de este [enlace](#).
- Amazon. (s.f.). *¿Cómo se calculan las calificaciones de estrellas del producto?* Recuperado de este [enlace](#).
- Amazon. (s.f.). Amazon Empire: The Rise and Reign of Jeff Bezos [Documental].
- Asch, S. E. (2018). *Opinions and social pressure*. Readings about the social animal (pp. 17-26). Worth Publishers.
- BBVA. (s.f.). *Aprinde a gestionar las devoluciones de Amazon*. Recuperado de este [enlace](#).
- Berkovsky, S., Kuflik, T., & Ricci, F. (2011). *The impact of data obfuscation on the accuracy of collaborative filtering*. Expert Systems with Applications
- Bowman, B. (s.f.). *Chapter 9 How to Master Amazon SEO and Move Your Products up the Search Rankings*. In *How to Sell on Amazon: The Ultimate (Step by Step) Beginner's Guide*. AMZ Profit Pros.
- Cañamares Pérez, R. (2019). *An analysis of popularity biases in recommender system evaluation and algorithms*. Universidad Autónoma de Madrid. Escuela Politécnica Superior Departamento de Ingeniería Informática.
- Chizari, N., Shoeibi, N., & Moreno-García, M. N. (2023). *A Comparative Analysis of Bias Amplification in Graph Neural Network Approaches for Recommender Systems*. Cornell University, ArXiv, electronics 2022, 11, 3301

- Chuah, G. S. F., Ngo, T. D., Pham, T. V., Lim, C. P., & Chua, T. S. (2021). *Reducing Exposure Bias in Recommender Systems: A Hybrid Approach*. IEEE Access, 9, 59514-59528.
- Deldjoo, Y., Bellogin, A., & Di Noia, T. (2021). *Explaining recommender systems fairness and accuracy through the lens of data characteristics*. Information Processing & Management, 58(2).
- Drucker P.F., (1954). *The practice of management*. New York: Harper & Row, New York, pp. 44-45.
- Flaxman, S., Goel, S., & Rao, J. M. (2016). *Filter bubbles, echo chambers, and online news consumption*. Harvard Public Opinion Quarterly, 80(Special Issue), 298–320.
- Fleder, D., & Hosanagar, K. (2010). *Recommender systems and their impact on sales diversity*. MIS Quarterly, 34(1), 37-53.
- Franklin, L. (2019). *Amazon vs. Google: Decoding the World's Largest E-commerce Search Engine*. Palgrave Macmillan.
- Garg, N. (2022). *Feature Engineering for Recommendation Systems – Part I*.
- Guimond, S., & Dambrun, M. (2018). *Social influence and conformity*. In *Advances in experimental social psychology* (Vol. 57, pp. 149-213). Academic Press.
- Hall, M., van der Maaten, L., Gustafson, L., Jones, M., & Adcock, A. (2022). *A Systematic Study of Bias Amplification*.
- Hanna, R. (2018). *Amazon SEO: The Ultimate Guide to Higher Product Ranking (With Tips & Examples)*.
- He, X., Liao, L., Zhang, H., Nie, L., Hu, X., & Chua, T. S. (2017). *Exposure Bias: A Hidden Challenge in Recommender Systems Research*. Proceedings of the 11th ACM Conference on Recommender Systems (pp. 23-31).
- Kumar, A., Gupta, M., Jain, V., & Varshneya, P. (2021). *Selection Bias in Recommender Systems: A Review*. IEEE Access, 9, 139878-139898.
- Madrigal, L. (2012). *The analysis of variance (ANOVA)*. In *Statistics for Anthropology* (pp. 122-145). Cambridge University Press. Recuperado el 5 de junio de 2012.
- Mansoury, M., Abdollahpouri, H., Pechenizkiy, M., Mobasher, B., & Burke, R. (2020). *Feedback Loop and Bias Amplification in Recommender Systems*. Proceedings of the 29th ACM International Conference on Information and Knowledge Management (CIKM'20), October 19–23, 2020, Ireland (pp. 1395-1399). ACM, New York, NY, USA.

- Moliner, J., Sainz, J., Pérez-González, F., & del-Pozo-Coronado, J. (2018). *Factors that influence users' intention to download mobile apps: A qualitative study*. Journal of Business Research, 85, 222-230.
- Moreno, L. (17 de febrero de 2023). *Diferencias estadísticas: z-test, t-test y ANOVA*. RPubS by RStudio.
- Peiró, K. & Baeza-Yates, R. (2021, 2 de marzo). *¿Puede la IA crear un mundo más justo?* Centro de Cultura Contemporánea de Barcelona.
- Peska, L. (2017). *Multimodal Implicit Feedback for Recommender Systems*. ITAT 2017 Proceedings (pp. 240-245).
- Pipino, L. L., Lee, Y. W., & Wang, R. Y. (2002). Data Quality Assessment. Communications of the ACM, 45(4).
- Sar Shalom, O., Berkovsky, S., Ronen, R., Ziklik, E., & Amihod, A. (2015). *Data Quality Matters in Recommender Systems*. ACM. Viena, Austria
- Shekhar, G. (2021, March 11). *Bias Detection in Machine Learning Models using Amazon SageMaker Clarify*.
- Simeone, O. (2018). *A Very Brief Introduction to Machine Learning With Applications to Communication Systems*.
- Stahl, B. C., Schroeder, D., & Rodrigues, R. (2023). *Ethics of Artificial Intelligence: Case Studies and Options for Addressing Ethical Challenges* (pp. 15). Springer.
- Stone, B. (2013). *The Everything Store: Jeff Bezos and the Age of Amazon*. Little, Brown and Company.
- Sun, T., Zhang, Y., Hu, X., Xie, X., & Zhang, Y. (2021). Exposure Bias and the Business Value of Recommendations. En Proceedings of the 15th ACM Conference on Recommender Systems (pp. 206-214).

IX. CÓDIGO UTILIZADO

Para llevar a cabo el análisis práctico de los datos se ha utilizado la herramienta de R y el siguiente código:

```
library(readxl)
library(tidyverse)
library(tidyquant)
library(ggplot2)
tabla <- read_excel("Documents/Documentos/ICADE/TFG/ANALYTICS/5.
datos/tabla.xlsx")
View(tabla)

ntabla <- tabla[complete.cases(tabla[, c("user_attr")]),]
nrow(ntabla)

resultado_chi <- chisq.test(ntabla$model_attr, ntabla$user_attr)
resultado_chi$observed
resultado_chi$expected
resultado_chi$statistic

tabla_contingencia <- table(ntabla$model_attr, ntabla$user_attr)
tabla_contingencia

frecuencias_relativas <- prop.table(tabla_contingencia)
frecuencias_relativas

mtable<-as.data.frame(table(frecuencias_relativas$model_attr,
frecuencias_relativas$user_attr))
colnames(mtable)<- c("attrib", "selection", "freq")
mtable

ggplot(mtable, aes(attrib, selection)) +
  geom_tile(aes(fill = freq), colour = "black") +
  scale_fill_gradient(low="white", high="steelblue")

ggplot(mtable, aes(x = attrib, y = selection, size = freq, color = freq)) +
  geom_point() +
  labs(x = "attrib", y = "selection", size = "freq") +
  scale_color_gradient(low = "yellow", high = "red") +
  theme_bw()

tabla_contingencia <- table(ntabla$user_attr, ntabla$rating)
tabla_contingencia

frecuencias_relativas <- prop.table(tabla_contingencia)
frecuencias_relativas

with(ntabla, Hist(rating, scale="frequency", breaks="Sturges",
col="turquoise"))
```

```

mycol<-c("pink", "blue")
xyplot(rating ~ rating | user_attr, groups=user_attr, type="p", pch=16,
       auto.key=list(border=TRUE), par.settings=simpleTheme(pch=16, col=mycol),
       scales=list(x=list(relation='same'), y=list(relation='same')), data=ntabla)
with(ntabla, Hist(rating, groups=model_attr, scale="frequency",
                 breaks="Sturges", col="turquoise"))

with(ntabla, plotMeans(rating, user_attr, error.bars="se", connect=TRUE))
with(ntabla, plotMeans(rating, model_attr, error.bars="se", connect=TRUE))

with(ntabla, Hist(rating, groups=category, scale="frequency",
                 breaks="Sturges", col="turquoise"))

AnovaModel<- aov(rating ~ user_attr, data=ntabla)
summary(AnovaModel)
with(ntabla, numSummary(rating, groups=user_attr, statistics=c("mean",
                                                             "sd")))

AnovaModel1<- aov(rating ~ model_attr, data=ntabla)
summary(AnovaModel1)
with(ntabla, numSummary(rating, groups=model_attr, statistics=c("mean",
                                                                "sd")))

AnovaModel2<- aov(rating ~ category, data=ntabla)
summary(AnovaModel2)
with(ntabla, numSummary(rating, groups=category, statistics=c("mean",
                                                             "sd")))

```