



UNIVERSIDAD DE CHILE

FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS

DEPARTAMENTO DE INGENIERÍA INDUSTRIAL

EVALUACIÓN DE LA EFECTIVIDAD DE UN SISTEMA DE RECOMENDACIÓN DE PRODUCTOS EN
UNA TIENDA POR DEPARTAMENTOS

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL

GUILLERMO HERNÁN RUBILAR FERIS

PROFESOR GUÍA:

PABLO MARÍN VICUÑA

MIEMBROS DE LA COMISIÓN:

ALEJANDRA PUENTE CHANDIA

JUAN PABLO ROMERO GODOY

SANTIAGO DE CHILE

2019

EVALUACIÓN DE LA EFECTIVIDAD DE UN SISTEMA DE RECOMENDACIÓN DE PRODUCTOS EN UNA TIENDA POR DEPARTAMENTOS

Los sistemas de recomendación son herramientas utilizadas por retailers para personalizar la experiencia de compra. Un ejemplo de esto es Alibaba, el cual señala que, en sus campañas promocionales, las ventas realizadas en páginas con recomendaciones personalizadas tienen una conversión un 20% mayor que las no personalizadas.

El trabajo de título consiste en evaluar el desempeño de metodologías para la recomendación de productos en las categorías de vestuario y calzado en una tienda por departamentos. El desarrollo tiene dos fases: Seleccionar, en base a la eficiencia y eficacia, el modelo de filtros colaborativos que entrega mejores resultados de acuerdo con la métrica del recall. Luego se desarrolla un A/B test en email-marketing para medir la efectividad empírica del modelo seleccionado con respecto a recomendaciones del área comercial del negocio.

De la primera fase se tiene que, con 5 recomendaciones por cliente, el mejor modelo de la familia de factorización de matrices (filtro colaborativo basado en modelo) acierta a alguna de las compras de un 29% de los clientes, mientras que la mejor implementación en base a KNN (filtro colaborativo basado en memoria) acierta a un 32%. Por contraparte KNN tiene un tiempo de ejecución 11 veces mayor, tiempo que además no aumenta de forma lineal al aumentar el número de clientes. Por lo tanto, se elige el de FM como el modelo a probar experimentalmente debido a que tiene mayores posibilidades para una implementación real.

Con respecto al experimento, se envía un correo con recomendaciones dadas por el modelo de factorización de matrices y otro con el mismo diseño y asunto, pero con recomendaciones masivas elegidas por el área comercial. Como resultados se tiene que las recomendaciones personalizadas generaron un 32% más de tráfico hacia la página (10.123 v/s 7.645), 69% más de conversión en órdenes (123 v/s 74) y un 7% más de venta en la página web incluyendo ventas cruzadas (3.411.760 CLP v/s 3.187.540 CLP).

Se concluye que, para el caso analizado, hacer recomendaciones personalizadas en base a factorización de matrices entrega mejores resultados que recomendaciones generales basadas en juicio de experto. Además, se proponen algunos puntos de contacto con el cliente donde podrían ser implementadas estas recomendaciones, como pueden ser: campañas de email, página web, selección de las categorías de descuentos para los cupones que se entregan en caja y selección de categorías para cupones que se envían al hacer una compra online con retiro en tienda. Finalmente, también se propone que la implementación de filtros colaborativos podría aportar a otros negocios del holding que tengan un alto volumen de sku para recomendar, en específico se cree que los de retail como supermercados, mejoramiento del hogar y marketplace serían buenos candidatos para realizar más pruebas sobre estas metodologías.

Tabla de contenido

1.	Introducción	1
2.	Descripción del proyecto	2
2.1.	Justificación	2
2.2.	Objetivos y alcance	3
2.2.1.	Objetivo general	3
2.2.2.	Objetivos específicos	3
2.2.3.	Alcance	4
3.	Marco teórico	5
3.1.	Filtros colaborativos	6
3.1.1.	Vecino más cercano (KNN) - Recomendación colaborativa basada en memoria	6
3.1.2.	Factorización de matrices– Recomendación colaborativa basada en modelos	7
3.1.3.	Ventajas y desventajas de los filtros colaborativos	8
3.2.	Filtros basados en contenido	9
3.2.1.	Árboles de decisión	9
3.2.2.	Regresión logística	10
3.3.	Recomendaciones basadas en conocimiento	11
3.4.	Recomendaciones híbridas	11
3.5.	A/B test	12
4.	Metodología	12
4.1.	Fase analítica	13
4.2.	Selección y pre-procesamiento de los datos	13
4.3.	Métricas para medir el desempeño de los modelos	14
4.4.	Metodología de implementación de modelos de factorización de matrices	15
4.5.	Metodología de implementación de KNN	16
4.6.	Fase de experimentación	16
5.	Desarrollo de la metodología	18
5.1.	Preparación de los datos	18
5.1.1.	Generación de la base de clientes y tabla de transacciones	18
5.1.2.	Estacionalidad en vestuario y calzado	21
5.2.	Modelos para establecer un punto de comparación	22
5.3.	Factorización de matrices	24
5.3.1.	Modelo base	24

5.3.1.1.	Calibración de alfa y número de factores a utilizar	24
5.3.1.2.	Análisis de resultados y escalabilidad del modelo	26
5.3.2.	Iteración 1: No recomendar clases-marcas ya compradas	28
5.3.2.1.	Análisis de resultados	28
5.3.3.	Iteración 2: Incluir la estacionalidad mensual de las clases	29
5.3.3.1.	Desarrollo de metodología	29
5.3.3.2.	Análisis de resultados	30
5.3.4.	Iteración 3: Añadir gasto pasado del cliente	31
5.3.4.1.	Desarrollo de la metodología	31
5.3.4.2.	Análisis de resultados	33
5.3.5.	Conclusiones factorización de matrices	34
5.4.	Recomendaciones de KNN	35
5.4.1.	Modelo base	35
5.4.1.1.	Análisis de resultados	36
5.4.2.	Iteración 1: Recomendar en base a lo que compran los vecinos por sobre las compras generales	37
5.4.3.	Iteración 2: No considerar transacciones con ofertas en la selección de la clase-marca	38
5.4.4.	Conclusiones KNN	39
5.5.	Selección de modelo a utilizar en fase experimental	40
6.	Desarrollo metodología fase de experimentación	41
6.1.	Preparación del envío	41
6.1.1.	Diseño de la gráfica	41
6.1.2.	Elección de clientes	43
6.1.3.	Obtención de recomendaciones	44
6.2.	Resultados del envío	46
7.	Conclusiones	51
8.	Bibliografía	52
9.	Anexos	53
9.1.	Anexo A: Matriz con compras de clientes	53
9.2.	Anexo B: Guess segmento muy alto	54

1. Introducción

El trabajo de título se desarrolla en una importante tienda por departamentos nacional, donde se considera como sus principales clientes a personas naturales. De acuerdo a la participación en ventas, el negocio se encuentra dentro de los primeros puestos de en el mercado nacional. Debido a que la compañía pertenece a un conglomerado de larga tradición se deben realizar cada vez mayores inversiones para seguir creciendo y adaptándose al entorno competitivo nacional e internacional, es por esto que, durante los últimos años el énfasis de la empresa ha estado en el crecimiento regional, la digitalización de las compañías y la automatización de los procesos. En particular en el plan de inversión al corto plazo se tiene un foco en las inversiones de logística y transformación digital. Este nuevo rumbo que están tomando los negocios se ve reflejado en la creación de plataformas de marketplace o en la adaptación a la omnicanalidad como uno de sus focos principales. Durante el año 2017, las ventas online de los negocios del retail fueron por sobre los US\$700 millones, lo que significa un 6.5% de las ventas totales del año en el ámbito del retail y alcanza un crecimiento de un 36% con respecto al año anterior. En la memoria anual de la empresa se destaca el objetivo de afianzar aún más la posición del negocio mediante las herramientas que abre el mundo digital, ya sea aumentando el surtido de producto disponible o personalizando las ofertas a los usuarios. Es en esta nueva búsqueda de oportunidades donde se enmarca el tema de memoria, el cual tiene como finalidad evaluar metodologías de recomendación de productos para activar consumidores del negocio de forma proactiva.

Actualmente en el negocio, los mecanismos más personalizados para activar clientes corresponden a las campañas de cupones y las de email marketing, donde destacan aquellas que se activan mediante triggers como las más específicas. Estas campañas de correos en el año 2016 generaron en torno a un 1% de los ingresos correspondientes a ventas identificables del negocio (en base a la venta incremental), de los cuales un 12.3% se señala que fue producto de los correos activados por uno de los siguientes tipos de trigger:

1. Venta cruzada: Se recomiendan productos en función a lo que ha comprado el cliente en los últimos días utilizando reglas de asociación.
2. Recompra: Algunos productos de perfumería tienen un periodo estimado de recompra. Frente a una compra antigua del usuario se utiliza este tiempo para activar el trigger.
3. Navegación: Es un correo con envío diario, se recomiendan los productos navegados el día anterior.

4. Carro abandonado: Es un correo con envío diario, se envía un recordatorio con los productos que están añadidos al carro de hace más de 24 horas.
5. Hitos: Se envían recomendaciones de productos en base a información sobre eventos en la vida del cliente, tales como: cumpleaños del cliente o de algún familiar, club novios y bebés.

De estas categorías de trigger se pueden ver dos características relevantes:

1. Se basan en la ocasión de consumo cliente y no en sus preferencias.
2. Todos los trigger, a excepción del activado por hitos y el de recompra, requieren que el cliente haya tenido algún tipo de interacción con el negocio recientemente (mediante compras o navegación). Un cliente que no ha interactuado con el negocio en un determinado tiempo no recibirá este tipo de comunicaciones personalizadas.

Otro punto importante a destacar es el crecimiento sostenido que han tenido las ventas online en la participación del total, donde entre el 2016 y el 2018 ha aumentado en un 50% con respecto al monto y un 100% en el número de transacciones totales, por lo que se hace cada vez más relevante apoyar el crecimiento de este canal apuntando a la personalización del cliente como una de las estrategias relevantes.

Frente a este escenario es que surge el planteamiento de explorar metodologías de recomendación de productos como un método para atraer clientes de manera personalizada, analizando el comportamiento de cada cliente en conjunto con el de la población total de clientes para ofrecer artículos de forma proactiva en lugar de reaccionar a las interacciones que tiene el cliente con el negocio.

2. Descripción del proyecto

2.1. Justificación

Los sistemas de recomendación están ampliamente aceptados como una forma eficaz de atraer clientes. Una muestra de ello es el estudio empírico realizado en el año 2010 titulado “Empirical analysis of the impact of recommender systems on sales” [1], donde se sostiene que estos motores de recomendación tienen diversos beneficios como: unas mayores ventas, más flexibilidad en la fijación precios y efectos positivos en las ventas cruzadas, moderado por el tiempo entre la compra que desencadena la recomendación y el contacto con el cliente. Por otro lado, la experiencia realizada en el año 2014 presente en la publicación “Impact of recommender systems on sales volume and diversity” [2]

muestra que el desarrollo de un filtro colaborativo en base a los productos comprados por los clientes en un retailer online, tuvo como resultado el aumento en un 35% de las ventas y un 25% en las visitas a los productos.

Sobre los sistemas de recomendación no solo hay estudios con fines académicos. Actualmente dos grandes retailers del mundo como lo son Amazon y Alibaba ya han implementado estos motores. De acuerdo con estimaciones realizadas por Mckinsey [3] un 35% de las ventas totales de Amazon se deben a sus recomendaciones mientras que Alibaba señaló que durante su evento anual de ventas el “11.11 global shopping festival” del año 2016 se generaron 6.7 millones de páginas con recomendaciones personalizadas, las cuales tuvieron una conversión un 20% mayor que las no personalizadas [4].

Existe evidencia que la implementación de un motor de recomendaciones en los puntos de contacto del cliente con la empresa puede entregar beneficios económicos aumentando la conversión, por lo que se hace necesario evaluar que metodologías se ajustan mejor a la realidad del negocio y sus clientes.

2.2. Objetivos y alcance

2.2.1. Objetivo general

Evaluar el desempeño de metodologías de filtros colaborativos en la recomendación de productos personalizados, utilizando una validación analítica y experimental sobre la efectividad y eficiencia de los modelos implementados.

2.2.2. Objetivos específicos

- En base a métricas de clasificación, seleccionar que metodología de filtros colaborativos se ajusta mejor a la recomendación de productos de vestuario y calzado.
- Desarrollar una prueba mediante email marketing para comprobar la mejora que genera su implementación con respecto a recomendaciones comerciales.

2.2.3. Alcance

En primer lugar, para poder definir el alcance, es necesario describir la agregación de productos existente en el negocio, la cual va desde las categorías generales que vende el negocio (vestuario, decohogar y electrohogar) hasta el SKU individual. A continuación, se presenta un diagrama que representa los niveles de agregación disponibles y una tabla que muestra el rango en el que se encuentra el número de productos distintos en cada nivel durante el año 2017, donde se ve el aumento a medida que baja el nivel de agregación, llegando hasta un total entre 500.000 y 1.000.000 de sku.

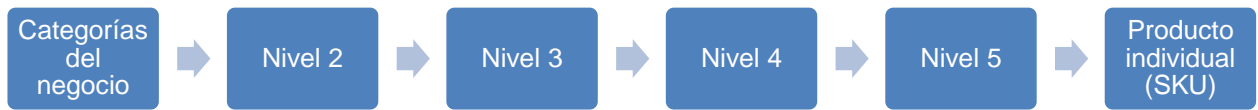


Diagrama 1-Desagregación de las categorías. Fuente: Elaboración propia

Agregación	Rango de elementos distintos en la agregación	Ejemplo
Nivel 1	3	Vestuario y calzado
Nivel 2	<100	HOMBRES
Nivel 3	[100-500]	TOPS SPORT
Nivel 4	[500-1000]	POLERAS
Nivel 5	[1000-3000]	POLERA ML LISA
Producto individual (SKU)	[500.000- 1MM]	5668082

Tabla 1-Número de elementos en cada nivel de agregación y ejemplo. Fuente: Elaboración propia

El trabajo de título tiene dos alcances con respecto a estas agrupaciones:

1. Las pruebas y recomendaciones a realizar se hacen con respecto a productos asociados al mismo género del cliente en vestuario y calzado.
2. Debido a la alta rotación y el gran volumen de SKU disponibles, la recomendación de productos se realiza de acuerdo al nivel de agrupación creado por la combinación del nivel 4 y la marca del producto (de aquí en adelante la combinación será llamada clase-marca). Utilizando la tabla 1 como referencia, un ejemplo sería “Polera-Adidas”.

En el caso de la experimentación, donde es necesario enviar un conjunto de sku’s en específico al cliente, el alcance limita esta elección al uso de reglas sencillas para seleccionarlos dentro de las clase-marca recomendadas.

3. Marco teórico

Los sistemas de recomendación pertenecen a la familia de los sistemas de filtrado de información, los cuales tienen por objetivo afrontar la sobrecarga de datos de la actualidad buscando utilizar la información más relevante para lograr un objetivo en concreto. En el caso de los motores de recomendación, esta información es generada desde las preferencias del usuario, sus intereses o desde el comportamiento observado sobre un producto [5]. El objetivo final de la recomendación consiste en predecir un comportamiento particular sobre si un cliente le agrada o no un producto basado en su perfil de compra.

Estos sistemas tienen consecuencias positivas tanto para los usuarios como para las tiendas. A los primeros les aporta en hacer más simple la búsqueda de algún producto en particular mientras que para las tiendas de e-commerce, han probado ser un medio efectivo para aumentar los ingresos en base a el aumento de las ventas de productos [5].

Todo sistema de recomendación debe generar un perfil o modelo del cliente, donde la forma en que la información es adquirida y usada depende directamente de la técnica de recomendación a utilizar. De acuerdo con la naturaleza de la obtención de la información del usuario ésta se puede separar en dos categorías:

- Información explícita: Recabada en base a las preferencias señaladas por el usuario y el feedback recibido, como, por ejemplo, preguntar sobre categorías preferidas, recibir calificación sobre productos consumidos (puede ser respuesta binaria o en base a ratings) o evaluación sobre la calidad de recomendaciones anteriormente realizadas, entre otros.
- Información implícita: Ésta se obtiene directamente del monitoreo en el comportamiento del usuario sin la necesidad de que él tenga que hacer una observación directa sobre su perfil de compra. Algunos ejemplos de este tipo de fuentes de información son: compras anteriores, visitas de productos, búsquedas en la página, frecuencia de visita a la web, etc.

Este perfil de usuario puede ser usado con información adicional como las características de los productos o la opinión de los demás usuarios sobre algún artículo en particular para realizar el proceso de recomendación. Es de esta combinación de información desde donde nacen cuatro tipos de aproximaciones para el desarrollo del sistema de recomendación: recomendaciones colaborativas, recomendaciones basadas en contenido, recomendaciones basadas en conocimiento y aproximaciones híbridas.

3.1. Filtros colaborativos

La idea básica de estos filtros es que si usuarios han compartido los mismos intereses en el pasado también lo harán en el futuro [6]. Por ejemplo, si hay dos usuarios A y B que han tenido un historial de compras muy parecido, entonces se podría recomendar a B productos que ha comprado A que B aún no ha comprado. Entre los principales métodos para desarrollar filtros colaborativos se tienen metodologías basadas en memoria, que utilizan la información de las valoraciones de productos disponibles para medir similitudes y recomendar los productos en base a las transacciones recientes de los similares y metodologías basadas en modelo que utilizan la misma fuente de información para intentar predecir cuanto le gustaría cada producto a un usuario. A continuación, se presentan el método de vecinos más cercanos y el método de factorización de matrices como ejemplo para cada uno de ellos.

3.1.1. Vecino más cercano (KNN) - Recomendación colaborativa basada en memoria

Este método busca realizar la recomendación en base a las valoraciones que usuarios históricamente han realizado sobre los productos, para esto existen dos enfoques con un funcionamiento similar:

1. Centrado en el usuario: Utiliza la similitud entre clientes, de modo que se pueda predecir la valoración que le daría un individuo a un ítem en base a las calificaciones que los demás le han otorgado anteriormente.
2. Centrado en el producto: Utiliza la similitud entre ítems, de modo que se pueda predecir la valoración que un usuario le dé a un producto en base a las calificaciones que ha otorgado con anterioridad a los demás.

Además, para la ejecución de los vecinos más cercanos se debe asumir dos supuestos:

1. Si un grupo de usuarios han tenido gustos similares en el pasado, entonces también los tendrán en el futuro
2. Las preferencias de los usuarios son consistentes a lo largo del tiempo

Entonces, explicando el método en base a una recomendación centrada en el usuario. Dado un conjunto de usuarios, un conjunto de productos y la matriz de ratings que corresponde a la nota que le pone un usuario a un producto (no todos los usuarios valoran todos los productos) y un individuo I para el cual hacer la recomendación, ésta se puede lograr en dos pasos:

1. Identificar clientes similares a I en función de la información disponible: Para esto se utilizan medidas de similitud.
2. Utilizar la similitud con los demás individuos y la matriz de ratings para seleccionar los productos preferidos por la vecindad para ser recomendados a I.

Para identificar la similitud entre clientes se proponen tres medidas distintas:

- Correlación de Pearson: Entre dos clientes X e Y que han valorado un conjunto de productos ordenados en los vectores VX y VY se puede calcular como:

$$\rho_{X,Y} = \frac{COV(VX, VY)}{Desvest(VX) * Desvest(VY)}$$

Este indicador es el que mejor se comporta en recomendaciones donde se tengan ratings explícitos para recomendaciones centradas en los clientes [7].

- Distancia coseno: Entre dos clientes X e Y que han valorado un conjunto de productos ordenados en los vectores VX y VY se puede calcular como:

$$Similitud\ coseno_{X,Y} = \frac{VX \cdot VY}{||VX|| * ||VY||} = \frac{\sum_{i=1}^N VX_i * VY_i}{\sqrt{\sum_{i=1}^N VX_i^2} * \sqrt{\sum_{i=1}^N VY_i^2}}$$

Esta medida al ser la que genera resultados más precisos en las recomendaciones centradas en el producto se ha establecido como la métrica estándar en esos casos [6].

- Índice de Jaccard: El coeficiente de Jaccard es utilizado cuando la matriz de valoraciones contiene valores binarios (*like/dislike* o *compra/ no compra*, por ejemplo) en lugar de ratings continuos u ordinales [8]. Para un par de clientes X e Y que han comprado (o dado valoración positiva) a un conjunto de productos CX y CY se calcula como:

$$Jaccard_{X,Y} = \frac{CX \cap CY}{CX \cup CY}$$

3.1.2. Factorización de matrices– Recomendación colaborativa basada en modelos

Mientras que los modelos de recomendación basados en vecindades se enfocan en encontrar relaciones entre usuarios o entre productos, los modelos de factorización de matrices proponen caracterizar la relación entre ellos (la matriz de ratings mencionada anteriormente) en base a factores latentes inferidos desde patrones ocultos en la relación [9]. Los valores de los factores representan la importancia de alguna característica (que no es conocida), su relación con los ítems refleja el grado de presencia de estas características en los productos y la relación con los usuarios muestra la afinidad que tiene cada uno con estos atributos. Formalmente estas relaciones se denotan mediante los siguientes vectores:

- P_u : Que contiene la información sobre la relación de un usuario u con respecto a los distintos factores latentes.
- Q_i : Que contiene la información sobre la relación de un producto i con respecto a los distintos factores latentes.

Luego la predicción sobre la valoración de un usuario u sobre un producto i , corresponde al producto punto entre ambos vectores.

3.1.3. Ventajas y desventajas de los filtros colaborativos

Entre las principales ventajas del filtrado colaborativo se encuentra que solo necesitan información histórica sobre las transacciones (o valoraciones) para hacer las recomendaciones. Lo cual tiene como efecto negativo que, al trabajar con grandes conjuntos de usuarios o productos a recomendar, la matriz de valoraciones sea escasa (sparse).

Otra ventaja importante que tienen los filtros colaborativos es que son capaces de recomendar productos sin la necesidad de conocer gustos específicos de los usuarios y así generar una mayor diversidad de recomendaciones.

El último problema por considerar es el de “*cold start*”, donde para un producto o un cliente nuevo, no se le pueden hacer recomendaciones o no serán recomendados (dependiendo el tipo de enfoque que se esté tomando) hasta que tengan una cantidad de información suficiente.

Finalmente, entre KNN y factorización de matrices, se tiene la escalabilidad como principal diferencia, donde el segundo método tiene una mayor economía al momento de realizar las recomendaciones con la desventaja de que puede llevar a peores resultados que KNN debido a la pérdida de información que conlleva el hacer la reducción de dimensionalidad.

3.2. Filtros basados en contenido

Los filtros basados en contenido se basan principalmente en la disponibilidad de descripciones de los productos y un perfil de usuarios que les asigne importancia a las distintas características descritas [6] para generar la mejor recomendación en base a las preferencias del cliente. Un ejemplo sencillo sería la recomendación de películas, donde si un usuario sostenidamente ha visto películas de acción el sistema le ofrecerá automáticamente películas dentro del mismo género. Entre las metodologías aplicables más famosas en base a contenido se encuentran los modelos TF-IDF que son ampliamente utilizados en la recomendación de documentos, libros o artículos escritos, por otro lado, en cuanto a la recomendación de productos, se tienen a los árboles de decisión [2] [11] y la regresión logística [5] como metodologías que, además, son ampliamente utilizadas en múltiples ámbitos de la ingeniería.

3.2.1. Árboles de decisión

La implementación de árboles de decisión en la recomendación de productos tiene por objetivo generar un modelo capaz de predecir si a un usuario le gustará o no un producto en base a los atributos de este, de modo que las características de cada *item* son utilizadas para explicar las preferencias del usuario. En este caso, en el árbol, cada nodo de decisión refleja la presencia o no de algún atributo en el objeto mientras que el nodo final corresponde a la predicción de que si le gustará o no el producto al cliente.

A continuación, se presenta un ejemplo de árbol simple de la recomendación de una película para un usuario:

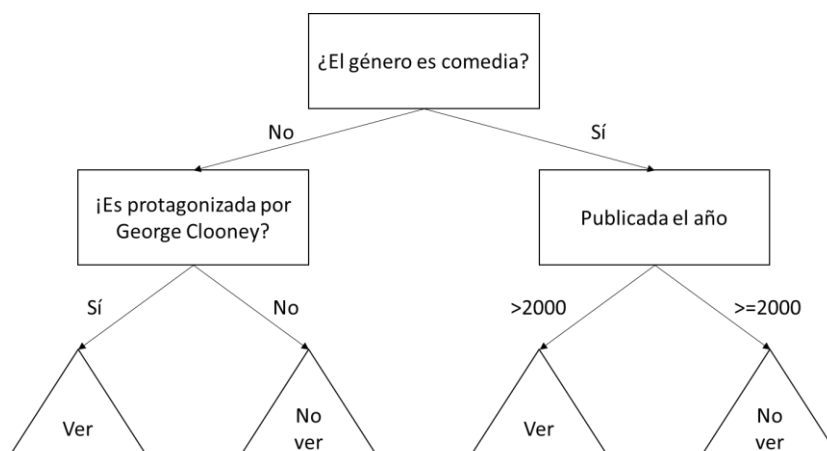


Diagrama 2- Ejemplo árbol de decisión para una película. Fuente: [11]

Entonces en la selección de los productos a recomendar, para cada usuario se deben valorar todos los productos según su árbol generado y así saber si le gustará o no el ítem.

Para aumentar la velocidad de ejecución en grandes volúmenes de clientes se propone crear los árboles a nivel de segmento de clientes de modo que las opciones a implementar quedarían como:

- Se puede generar un modelo para cada cliente y luego evaluar cómo sería el resultado de la recomendación de un producto para él.
- Sacrificar parte de la personalización de las recomendaciones y realizarlas a nivel de segmento de clientes, donde luego la valoración de la recomendación de cada producto sería de forma individual.

3.2.2. Regresión logística

El uso de una regresión logística simple en el ámbito de las recomendaciones se basa en que, de acuerdo con los atributos de un producto, es posible obtener una aproximación sobre cuánto le gustará a un usuario, por lo tanto, es necesario crear un modelo por cliente. En este caso las características de los objetos corresponden a las variables explicativas y el “cuanto” le gustaría el ítem al usuario pasaría a ser la variable dependiente. A continuación, se presenta como quedaría la regresión en la cual solo se utilicen atributos binarios como variable independiente, aunque también es posible agregar algunas continuas como el precio.

$$G(u, i) = \beta_0 + \sum_A \beta_{\text{atributo}} * x_j$$

Donde:

$G(u, i)$ es la función de utilidad del usuario u asociada al objeto i

$A(i)$ es el conjunto de atributos totales en el pool de productos

x_j es una variable Dummy, que representa la existencia o no del atributo j en un producto

Para hacer la predicción de afinidad entre usuario e ítem, solo basta reemplazar las variables independientes con 1 o 0 según corresponda, que corresponde a realizar la

cantidad de (Número de usuarios*Numero de ítems) valoraciones. Para grandes volúmenes de datos puede provocar graves ineficiencias en el tiempo de ejecución del modelo. Una forma de solucionar este inconveniente consiste en realizar segmentos de clientes en lugar de clientes en particular. El objetivo es disminuir este tiempo, sacrificando granularidad de personalización a favor de la eficiencia.

3.2.3. Ventajas y desventajas de los filtros en base a contenido

La principal ventaja de los filtros en base a contenido es no tienen problemas de “*cold start*” ya que generado un perfil de clientes posible recomendarle los productos de manera directa, sin que estos hayan tenido que tener interacciones con clientes.

Entre sus desventajas se tiene:

- Necesitar la información detallada de los usuarios y/o productos, para poder calibrar los modelos.
- Sobre-especialización en las recomendaciones, donde se genera una tendencia a recomendar los mismos productos al perfil del cliente.
- La necesidad de desarrollar un modelo por individuo en algunas de las metodologías hace costosa la ejecución para la calibración y predicciones.

3.3. Recomendaciones basadas en conocimiento

En los casos que la naturaleza de los productos involucre la venta a una gran proporción de “*one time buyers*” (como pueden ser automóviles, por ejemplo), no se puede confiar en las transacciones históricas como medio para generar perfiles de usuarios o sistemas colaborativos (basados en el cliente) para la recomendación. Es por ello que se hace necesario obtener información adicional del usuario, lo que se logra generalmente consultando de forma directa al cliente sobre los requerimientos mínimos del producto o mediante un diálogo interactivo para disminuir la barrera de conocimientos técnicos del producto.

3.4. Recomendaciones híbridas

Con el objetivo de enfrentar las ventajas y desventajas que puede tener las distintas herramientas para la recomendación es que se proponen las metodologías híbridas, que mediante la combinación de dos o más métodos de recomendación y/o reglas de filtrado, buscan lograr mejores resultados (en cuanto a eficiencia, eficacia o ambos) que las alternativas puras. Estas recomendaciones híbridas pueden ser realizadas en forma secuencial de modo que se pueda ir refinando la recomendación y que cada paso anterior

sea un insumo del siguiente o en paralelo y ponderar los resultados de los distintos modelos.

3.5. A/B test

Los A/B test corresponden a uno de los diseños experimentales más utilizados en marketing digital. En estos se busca que clientes distintos prueben distintas condiciones de un sitio web, e-mail, buscadores, ente otros. Para esto, los usuarios son asignados en un grupo de control y un grupo de tratamiento, donde el grupo de control en general corresponde a una versión actual del sistema y el grupo tratamiento incluye la variable sobre la que se quiere medir como afecta alguna métrica de desempeño. En el caso de sistemas de recomendaciones se utiliza para probar la efectividad de un modelo de recomendaciones con respecto a la situación actual [7] o también en varios modelos de manera simultánea (A/B/n test).

Para la separación de los grupos en un A/B test, una muestra lo suficientemente grande puede ser dividida de forma aleatoria asumiendo que no hay sesgo al clasificar los clientes. También, en caso de ser necesario es posible balancear ambos grupos en base a características conocidas que podrían influir en los resultados del experimento. Un método para hacer este balance consiste en el re-muestreo aleatorio, que corresponde a realizar iteraciones consecutivas, de modo que se siguen los siguientes pasos:

1. Separar de forma aleatoria el conjunto de clientes en dos grupos
2. Comparar si son similares en base a las características conocida
3. Si tiene mayor similitud que la mejor iteración realizada anteriormente, esta toma su lugar y se sigue iterando hasta cumplir con el número definido inicialmente.

4. Metodología

Para poder medir la eficiencia y eficacia de los modelos de recomendación en la empresa, se desarrolla una metodología que consta de dos macro etapas, la primera corresponde a una fase analítica, donde se prueban distintos métodos de recomendación para seleccionar el con mejor desempeño, mientras que la segunda etapa corresponde a una fase experimental, donde se comprueba si efectivamente las recomendaciones personalizadas entregan mejores resultados que las comunicaciones actualmente utilizadas por el negocio.

4.1. Fase analítica

En base a lo planteado en el marco teórico, se eligen los filtros colaborativos como principal estrategia para evaluar las recomendaciones personalizadas, esto es debido a las características del negocio: gran volumen de sku, de clientes y la ausencia de información estructurada sobre las características de los productos. Además, dado que no hay información explícita sobre las preferencias de los clientes, se utilizan variantes de los modelos que permiten integrar información implícita, específicamente se utilizan las compras de los clientes.

En cuanto KNN se elige la distancia Jaccard como medida de semejanza entre clientes, mientras que, para la implementación del modelo de factorización de matrices con información implícita, se utiliza el modelo desarrollado en el paper *“Collaborative Filtering for Implicit Feedback Datasets”* [11].

Para probar la eficacia y eficiencia de los modelos por desarrollar se utiliza un periodo de entrenamiento comprendido entre el 01/04/2017 al 31/03/2018, luego se analizan los aciertos de las recomendaciones de productos en el mes de abril del 2018.

4.2. Selección y pre-procesamiento de los datos

En la selección de datos se elige una muestra aleatoria de aproximadamente 100.000 clientes, los que durante el periodo de prueba deben tener un mínimo de compras de distintas clase-marca realizadas en categorías correspondientes al mismo género del cliente. Definir el mínimo de compras nace de la necesidad de tener suficiente información transaccional para cada usuario, ya que son un insumo crucial para entrenar los modelos, además este mínimo se debe ajustar a la realidad del negocio, donde el número de clase-marca distintos comprados en un año por cliente es bajo.

La selección del mínimo de clase-marca distintos se realiza mediante un análisis exploratorio sobre la cantidad de transacciones anuales por cliente; en base a estos resultados se utilizará el máximo entre el promedio y la mediana como punto de transacciones mínimas.

La etapa de pre-procesamiento de los datos, que es común para todos los modelos, corresponde a la generación de una tabla que contiene las transacciones a nivel de clase-marca por cliente, con el objetivo de funcionar como insumo para las implementaciones.

Cliente	Clase-marca	Transacciones
X	A	5
Y	B	3
Z	A	4
X	B	1

Tabla 2- Ejemplo transacciones agrupadas. Fuente: Elaboración propia

En la etapa de transformación, para los datos de entrenamiento del modelo de factorización de matrices es necesario crear una matriz de datos que represente las transacciones en base a los usuarios y las clase-marca.

Transacciones	A	B
X	5	1
Y		3
Z	4	

Tabla 3- Ejemplo transformación para factorización de matrices. Fuente: Elaboración propia

4.3. Métricas para medir el desempeño de los modelos

Antes de explicitar la metodología de desarrollo de los modelos, es necesario definir las métricas que medirán el desempeño de cada uno de ellos. Para esto se utilizan dos para eficiencia y dos para cuantificar la eficacia.

Las métricas de eficacia corresponden a:

a) Recall a k (R@k):

$$R@k = \frac{\# \text{ De recomendaciones acertadas dentro de las } k \text{ recomendadas a cada cliente}}{\# \text{ De transacciones totales realizadas en el periodo de pruebas por la muestra}}$$

b) Recall de clientes a k (RC@k):

$$RC@k = \frac{\# \text{ De clientes a los que se les acierta en al menos una de las } k \text{ recomendaciones}}{\# \text{ De clientes que compran en la muestra}}$$

Por otro lado, para comparar la eficiencia entre modelos se revisa:

- El tiempo que demoran en obtener las recomendaciones personalizadas.

- La capacidad de escalar a medida que aumenta el número de clientes.

Actualmente en el negocio no existe una metodología analítica en la recomendación de productos centrados en el cliente, por lo que, para poder medir el desempeño general de las metodologías a evaluar, se utiliza un modelo de recomendaciones básico como punto de comparación. Para esto se hace una implementación sencilla, seleccionando los 5 productos más vendidos por género durante el mes de marzo 2018 que serán comparados en base al R@5 y RC@5 en abril del 2018.

4.4. Metodología de implementación de modelos de factorización de matrices

Lo primero es implementar el modelo implícito de factorización de matrices en base a las compras pasadas del cliente, el cual permite recomendar productos ya comprados anteriormente.

Luego, para factorización de matrices implícita se proponen tres posibles modificaciones, las cuales se espera que tengan un efecto en los resultados del modelo:

- a) No permitir recomendar productos ya comprados anteriormente, bajo el supuesto que el cliente ya no tenga interés en comprar dos veces el mismo artículo.
- b) Agregar estacionalidad al modelo, ya que factorización de matrices no incorpora ningún tipo de temporalidad en sus inputs.
- c) Apoyar las recomendaciones utilizando información sobre el gasto pasado del cliente, con el objetivo de disminuir el predominio de los productos de consumo masivo en las recomendaciones de los filtros colaborativos.

La forma de iterar estos modelos es secuencial, donde se parte de un modelo actual que en un inicio corresponde al modelo base, luego:

1. Se prueba el punto (a), si esta iteración da mejores resultados que el modelo base, el modelo actual es reemplazado por esta variación.
2. Se prueba la lógica del punto (b) sobre el modelo actual, si esta iteración da mejores resultados que el modelo actual, este último es reemplazado por la nueva variación que incorpora estacionalidad en las recomendaciones.
3. Finalmente se prueba la lógica del punto (c) sobre el modelo actual, para dejar como el mejor modelo de factorización de matrices el que entregue mejores resultados entre el modelo actual y la iteración que incorpora gasto pasado del cliente en las recomendaciones.

4.5. Metodología de implementación de KNN

Para KNN la primera implementación a realizar corresponde a un modelo que utiliza la distancia Jaccard en base a las compras de clase-marca distintas para encontrar una vecindad de 500 clientes, posteriormente recomienda lo más comprado por los vecinos en el mes anterior.

En KNN las modificaciones al modelo pueden realizarse en dos ocasiones:

- a) Al elegir la vecindad, donde se puede modificar la métrica de distancia o las variables con la cual se elegirán a los vecinos (no tienen que ser transacciones necesariamente).
- b) Al recomendar los productos a los clientes dado un conjunto de vecinos.

En el caso de este proyecto se realizan iteraciones en las recomendaciones de los productos, ya que de acuerdo a la bibliografía la distancia Jaccard es la métrica que mejor se comporta para valores binarios (compra- no compra) y utilizar otras variables como sociodemográficas está fuera del alcance del proyecto.

Para cambiar la recomendación de productos se proponen dos iteraciones distintas:

- a) Recomendar productos que la vecindad valore más que el promedio de los clientes totales (Lift).
- b) No considerar compras que hayan sido influenciadas por ofertas, ya que pueden no reflejar correctamente el gusto del cliente sobre los productos.

Estas iteraciones, a diferencia de la sección de factorización de matrices, son excluyentes entre sí y por lo tanto se realizan tres pruebas independientes de KNN para elegir el que tiene mejor desempeño.

Finalmente, la fase analítica termina concluyendo entre los mejores modelos de factorización de matrices y KNN.

4.6. Fase de experimentación

Teniendo en consideración que el objetivo final de un sistema de recomendaciones es encontrar productos que sean del gusto del cliente, los resultados de la fase analítica solo pueden tomarse como una aproximación del desempeño de las recomendaciones (por ejemplo, puede haber productos recomendados que el cliente no vio en ningún momento). Por esto, es necesario desarrollar la fase experimental para ver el

comportamiento de los clientes frente a una situación en que los productos le son ofrecidos directamente.

El diseño experimental consiste en un A/B test con el objetivo de probar el modelo seleccionado en la fase analítica. Para poder realizarlo se utilizan dos grupos de clientes a los cuales se les envían dos correos de igual formato y asunto, pero con distinto contenido. En este caso particular, no se utiliza un tercer grupo sin envíos ya que se solo se busca medir el impacto de la personalización y no la ganancia “neta” de enviar un correo personalizado. Entonces, los grupos de control y tratamiento quedan de la siguiente forma:

- Grupo de control: Clientes a los que se les enviarán recomendaciones masivas en base a “apuestas” solicitadas a los product manager el área comercial para representar la situación actual (lógica de juicio de expertos).
- Grupo de tratamiento: Clientes a los que se les enviarán recomendaciones personalizadas en base al modelo de recomendación.

Para este test, se utiliza un subconjunto de clientes que cumplen con las siguientes condiciones:

- Tienen una tasa de apertura mayor a un 10% de los correos.
- Cumplen con el mínimo de transacciones definido en la fase analítica.
- Han abierto un correo de la empresa durante el 2019 para asegurar que el medio de contacto sigue vigente.

Luego los grupos de clientes son conformados utilizando el método de random resampling, para asegurar que ambos tengan el mismo gasto promedio, mismas compras promedio, tasa de aperturas y tasa de clics.

Los procesos de generación de las plantillas dinámicas de los productos y el envío del correo se hacen en coordinación con el área de Inteligencia web de la empresa, donde se debe agendar la fecha de envío de modo que se cumpla con la política de toques del negocio.

Los resultados del experimento se analizan en base a tres métricas distintas:

- *Click through rate (CTR)*: Mide que porcentaje de los productos enviados en el correo les llama la atención a los clientes, utilizando el número de clics que hace cada cliente en el html.

$$CTR = \frac{\# \text{ Clicks en el correo del grupo}}{\# \text{ De clientes que abren el correo en el grupo}}$$

- La tasa de conversión (TC) de los clientes: Busca medir cuantas compras puede generar el sistema de recomendación por sobre las recomendaciones masivas realizadas por el área comercial.

$$TC = \frac{\# \text{ Productos comprados por los clientes en el grupo}}{\# \text{ De clientes que abren el correo en el grupo}}$$

- El ingreso promedio por cliente (ARPU): Mide el ingreso promedio por cliente contactado que generan los correos.

$$ARPU = \frac{\text{Suma del gasto de los clientes que ingresan a la página en el grupo}}{\# \text{ De clientes que abren el correo en el grupo}}$$

Para poder medir diferencias reales de gasto entre los grupos, se propone un test de Lenvene de dos colas para asegurar igualdad de varianza y un test t de diferencia de medias de dos colas para asegurar que las medias son significativamente distintas. Para medir si hay diferencias significativas en cuanto a las tasas, en primer lugar, se utiliza un test Z de diferencia de proporción con dos colas para muestras independientes para verificar que son distintas y luego uno de una cola para validar que uno sea efectivamente mayor que el otro.

5. Desarrollo de la metodología

5.1. Preparación de los datos

5.1.1. Generación de la base de clientes y tabla de transacciones

De acuerdo con la información disponible en la base transaccional, durante el período comprendido entre el 01-04-2017 y el 31-03-2018 el número total clientes con alguna interacción identificada por rut, ya sea una compra o devolución en las líneas de vestuario y calzado es superior a los 6 millones. Este número de consumidores contiene ruts que no corresponde analizar como los que:

- Solo han realizado interacciones en tiendas que no corresponden a el negocio principal con el que se está trabajando, como pueden ser las “*stand alone*”.
- Solo han tenido transacciones de devoluciones y no ventas.
- También contiene ruts considerados como no válidos, tales como los menores a 10.000 o que estén presentes en la lista de los considerados como ruido operacional, que corresponden a números genéricos que ingresan los cajeros (como por ejemplo 555.555).

Entonces, limpiando de la base de datos las transacciones que cumplen alguna de esas condiciones, queda un total superior a 5.000.000 de clientes únicos, de entre ellos, los

que tienen compras en líneas y sublíneas de vestuario y calzado para su mismo género (ya que estas transacciones serán utilizadas para calibrar los modelos) están en un rango comprendido entre 3.000.000 y 4.000.000 distribuidos de modo que un 64% corresponden al género femenino y 36% al masculino.

Luego de tener la cantidad de clientes válidos que compran categorías del mismo género, es posible ver la distribución de clientes de acuerdo con el número de clase-marca que compran en el periodo de entrenamiento.

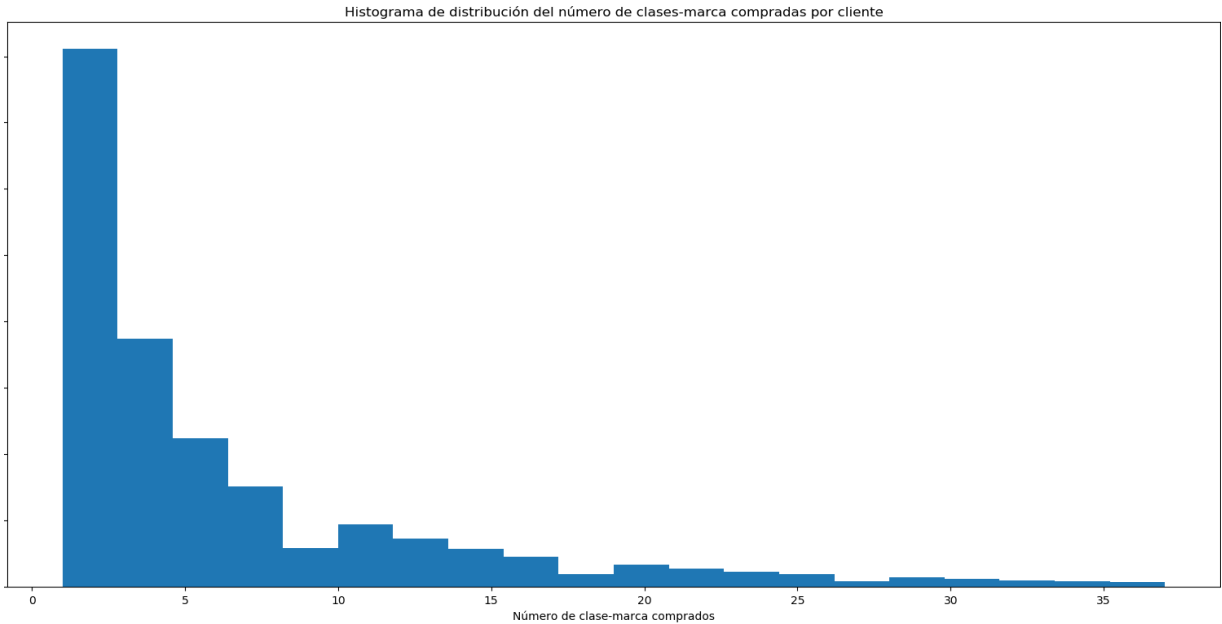


Gráfico 1- Distribución de clase-marca, valores eje Y confidenciales. Fuente: Elaboración propia

De la distribución se ve que, en general, los clientes tienen un bajo número de compras de clase-marca distintos en el año, donde el promedio corresponde a 4.55 y la mediana está en torno a 3. Para que los modelos sean capaces de detectar patrones correctamente, se eligen clientes que tengan como mínimo 5 clase-marca distintos comprados en el año y un máximo de 35, para evitar que outliers influyeran los resultados. Finalmente se genera una muestra aleatoria en torno a los 100.000 clientes que cumplen la condición señalada, donde la distribución de género y clase-marca se muestran a continuación.

Género	Porcentaje en el grupo total	# Clientes de la muestra	Porcentaje en el grupo de muestra
FEMENINO	78%	75.498	76%
MASCULINO	22%	24.465	24%

Tabla 4- Distribución de género en la muestra. Fuente: Elaboración propia

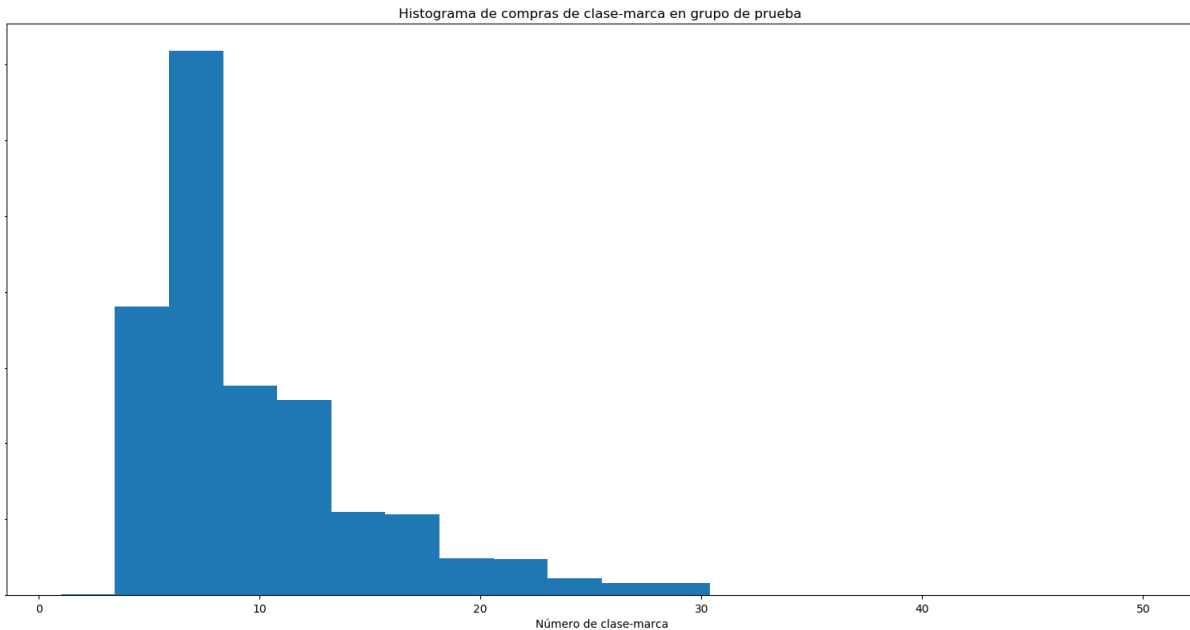


Gráfico 2-Distribución de clase-marca en la muestra, valores eje Y confidenciales. Fuente: Elaboración propia

El último de los preparativos necesario para desarrollar los modelos es obtener las transacciones agrupadas a nivel de clase-marca para cada cliente durante el periodo de entrenamiento, en la tabla 5 se puede ver un ejemplo donde: `mask_cliente` es un identificador único para cada cliente (el rut real es enmascarado), `clase` corresponde a la clase-marca y `cantidad` a la suma de unidades que el cliente compró. El número de clase-marca distintas disponibles en el conjunto de entrenamiento se encuentra en los rangos entre [500-1000] para hombres y [1500-2000] para mujeres mientras que el número de interacciones de clientes con clase-marcas distintas están entre [150.000-200.000] y [700.000-800.000] respectivamente. Un ejemplo de la forma matricial que tendría la tabla se puede ver en anexo A que muestra las transacciones realizadas por clientes de la tabla 5, donde se puede ver la gran predominancia de espacios en blanco que representan productos que los clientes no han comprado.

mask_cliente	clase	cantidad
12822	ZAPATOS HOMRE-MP (1)	2
14181	PANTALON HOMBRE-K2	1
14187	JEANS MODA-MP (2)	2
13598	CAMISA ML-VANDINE	1
5662	PANTALON-MP (3)	1
13591	CASACA CUERO- MP (1)	1
11190	PANTALÓN-DOCKERS	4
14232	SWEATER SPORT-MP (4)	1

Tabla 5- Transacciones agrupadas a nivel clase-marca por cliente, en esta gráfica confianza se refiere al número de transacciones de la clase marca en el periodo de entrenamiento y MP(n) reemplaza el nombre de alguna marca propia de la tienda

5.1.2. Estacionalidad en vestuario y calzado

Uno de los puntos importantes a analizar antes de desarrollar los modelos de factorización de matrices, corresponde a verificar la existencia de estacionalidad en la venta de productos de vestuario y calzado, para esto se realizan los siguientes pasos:

- 1) Obtener las ventas de cada mes y clase para los años 2016 y 2017
- 2) Calcular las ventas promedio mensual para cada una de las clases utilizando los 24 meses.
- 3) Para cada clase y cada mes, calcular el promedio de ventas usando los años 2016 y 2017.
- 4) Finalmente, en cada combinación de mes y clase se calcula:

$$\text{Proporción con respecto a la media anual}_{c,m} = \frac{\text{Promedio}(\text{ventas 2016}, \text{ventas 2017})_{c,m}}{\text{Promedio mensual (2016, 2017)}_c}$$

El indicador refleja que tan sobre o bajo la media está la clase cada mes del año, por lo tanto, mientras más desviación estándar haya entre los meses se propone que las ventas son menos estables durante el año, caso contrario implica que las ventas de cada mes están más cerca del promedio anual. En la tabla 6 se presentan como ejemplo las 5 clases con más y menos desviación estándar mensual.

DESCRIPCION CLASE	DEVEST
ZAP AGUA HOM	2,21
TRAJES DE BANO	1,76
SHORT	1,76
TRAJE DE BANO	1,68
JEANS SHORT	1,67
ZAP RUNNING HOM	0,18
ZAPATOS HOMBRE	0,18
PANTALONES	0,18
PANTALON CONFECCION	0,17
ZAP OUTDOOR HOM	0,15

Tabla 6- Clases-marca con más y menos estacionalidad. Fuente: elaboración propia

En los gráficos 3 y 4 se puede ver cómo se comporta en los últimos 3 años la proporción con respecto a la media de dos clases pertenecientes al grupo con más y menos desviación estándar. Se puede ver que la clase llamada “pantalón de confección” varía su indicador en rangos cercanos uno, mientras que en el gráfico 4, se ve que el indicador durante el año de la clase “Traje de baño hombre” es muy cercano a cero entre los meses de abril y septiembre para luego tener una gran alza durante los meses de verano, lo que es esperable y acorde a la intuición.

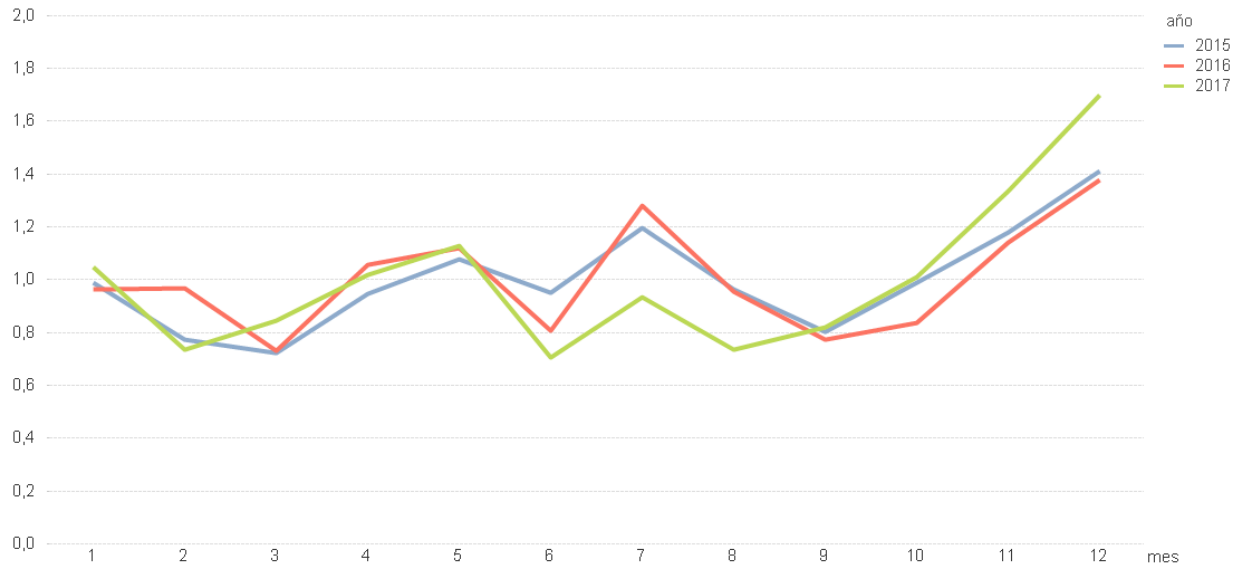


Gráfico 3- Estacionalidad "pantalón de confección". Fuente: elaboración propia

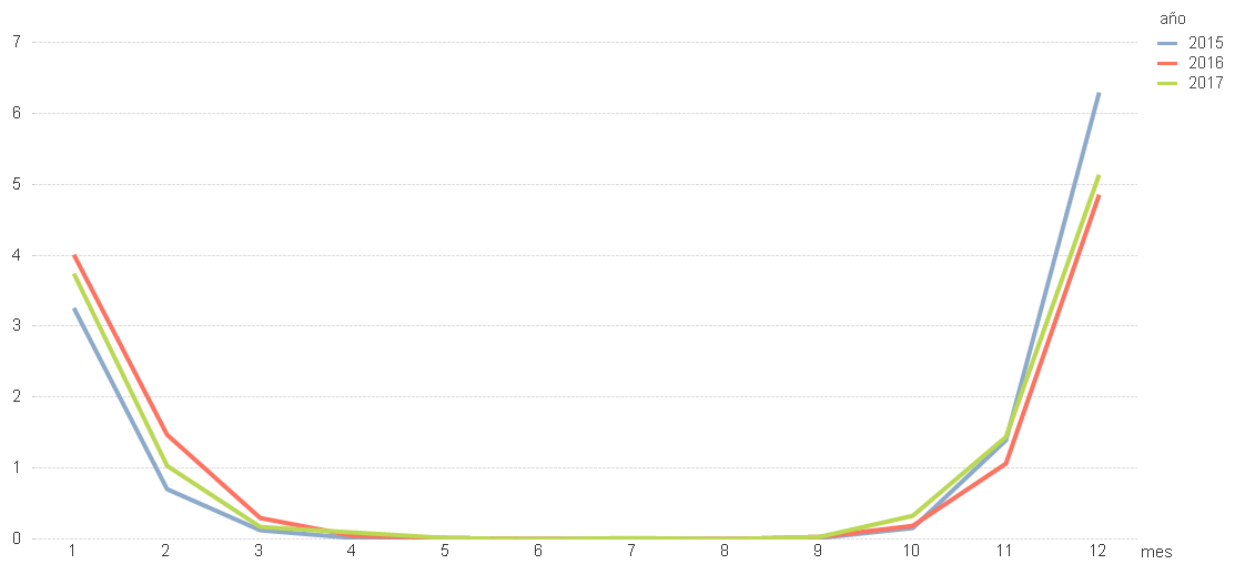


Gráfico 4- Estacionalidad "traje de baño hombre". Fuente: Elaboración propia

5.2. Modelos para establecer un punto de comparación

Con el objetivo de tener un punto de comparación para medir la efectividad de los modelos a desarrollar, es que se genera la aproximación basada en lo más vendido, donde se recomiendan las 5 clase-marca más vendidas durante marzo del 2018 a personas del mismo género.

Para el caso del género femenino, en la tabla 7 se muestran las clase-marca elegidas dentro del total, representando un 17% de las ventas en líneas de vestuario y calzado de mujer.

Clase marca	Posición	Porcentaje del total
JEANS DENIM-MARCA PROPIA 1	1	8%
POLERA ESTAMPADA-MARCA PROPIA 1	2	3%
BLUSA CONFECCION-MARCA PROPIA 2	3	2%
SWEATERS-MARCA PROPIA 1	4	2%
POLERA DISENO-MARCA PROPIA 2	5	2%
TOTAL	-	17%

Tabla 7- Clase-marca seleccionadas para mujer. Fuente: Elaboración propia

En la tabla 8 se muestran las recomendaciones para el género masculino, que representan un 12% del total de transacciones realizadas por hombres en categorías masculinas de vestuario y calzado.

Clase marca	Posición	Porcentaje del total
POLERA MC-MARCA PROPIA 3	1	3%
JEANS MODA-MARCA PROPIA 3	2	3%
PANTALON-MARCA PROPIA 4	3	3%
CAMISA SP ML-MARCA PROPIA 5	4	2%
POLERAS-MARCA PROPIA 5	5	2%
TOTAL	-	12%

Tabla 8- Clase-marca seleccionadas para hombres. Fuente: Elaboración propia

Con las recomendaciones ya obtenidas se procede a evaluar la efectividad al recomendar los productos seleccionados a la muestra de 99.963 clientes en abril del 2018, los resultados se pueden ver en la tabla 9, donde 26.860 clientes de los 99.063 realizaron 59.554 transacciones en vestuario y calzado en líneas del mismo género.

K	Aciertos	Aciertos@k	Recall@k	Aciertos clientes @k	Recall clientes @k
1	1962	1962	3%	2.007	7%
2	1480	3442	6%	3.399	13%
3	853	4295	7%	4.025	15%
4	947	5242	9%	4.637	17%
5	1624	6866	12%	5.678	21%

Tabla 9- Resultados más comprados mes anterior. Fuente: Elaboración propia

De los resultados en la tabla 9, se tiene que el benchmark para medir la eficacia de los modelos en los siguientes capítulos queda definido como $RC@5=21\%$, el cual señala que es capaz de acertar a alguna de las compras de un 21% de los clientes en la muestra. Además, una observación interesante sobre esta implementación, es que siendo k la

posición en el ranking que tiene el producto que se está recomendando al cliente (dependiendo del género), tener más transacciones en marzo 2018 no implica que en abril el producto tenga más aciertos que otros con peor ranking, esto es visible por ejemplo si comparamos los aciertos de $K=3$ y $K=5$, donde esta últimas tuvieron un 90% más de aciertos en el mes de test.

5.3. Factorización de matrices

Como se menciona en la metodología, la implementación del modelo de factorización de matrices se realiza en Python, utilizando el paquete “*implicit*” [12]. Con el código implementado es posible obtener recomendaciones de clase-marca ordenadas para cada cliente. En las siguientes secciones se muestran los resultados de las variantes de modelo realizadas sobre esta metodología.

5.3.1. Modelo base

5.3.1.1. Calibración de alfa y número de factores a utilizar

Utilizando las tablas de transacciones agrupadas como insumo para el modelo, y la condición de que se pueda recomendar al cliente una clase-marca ya comprada anteriormente, es que se realiza una *búsqueda en la grilla* para determinar el parámetro alfa y número de factores bajo los cuales este modelo entrega los mejores resultados en base al $R@5$ de la muestra para abril 2018.

Etiquetas de fila	1	2	3	4	5	6	7	8	9	10
1	12,71%	12,71%	12,71%	12,71%	12,71%	12,71%	12,71%	12,71%	12,71%	12,71%
2	13,73%	14,11%	13,91%	13,85%	13,79%	13,69%	13,68%	13,65%	13,61%	13,69%
3	14,99%	15,16%	15,20%	15,20%	15,12%	15,08%	15,04%	14,98%	14,93%	14,90%
4	15,01%	15,12%	15,20%	15,27%	15,22%	15,14%	15,00%	14,96%	14,87%	14,84%
5	14,87%	15,46%	15,50%	15,46%	15,43%	15,28%	15,08%	14,86%	14,91%	14,78%
6	15,16%	15,57%	15,63%	15,58%	15,50%	15,34%	15,32%	15,11%	15,06%	14,52%
7	15,21%	15,53%	15,35%	15,36%	15,41%	15,26%	15,35%	14,84%	15,09%	14,81%
8	15,17%	15,41%	15,42%	15,28%	15,20%	15,03%	14,88%	14,96%	14,65%	14,50%
9	15,08%	15,38%	15,29%	15,35%	15,27%	15,07%	14,92%	14,81%	14,70%	14,49%
10	15,23%	15,48%	15,34%	15,32%	15,34%	15,02%	14,97%	14,74%	14,56%	14,52%
11	15,08%	15,46%	15,56%	15,48%	15,11%	15,17%	14,96%	14,81%	14,61%	14,35%
12	15,00%	15,39%	15,58%	15,44%	15,44%	15,13%	15,05%	14,82%	14,55%	14,44%
13	15,01%	15,54%	15,70%	15,53%	15,32%	15,19%	14,99%	14,94%	14,72%	14,38%
14	14,97%	15,49%	15,53%	15,45%	15,45%	15,12%	15,00%	14,86%	14,70%	14,55%
15	15,08%	15,62%	15,53%	15,48%	15,38%	15,25%	15,12%	14,85%	14,64%	14,57%
16	15,10%	15,53%	15,53%	15,35%	15,22%	15,10%	15,00%	14,97%	14,60%	14,59%
17	15,14%	15,46%	15,50%	15,40%	15,36%	15,17%	15,07%	14,83%	14,69%	14,37%
18	15,05%	15,43%	15,53%	15,29%	15,31%	15,13%	14,94%	14,85%	14,69%	14,45%
19	15,09%	15,41%	15,43%	15,36%	15,37%	15,11%	14,98%	14,77%	14,49%	14,48%
20	15,14%	15,44%	15,45%	15,31%	15,15%	15,13%	14,91%	14,73%	14,56%	14,37%
21	15,10%	15,30%	15,35%	15,37%	15,27%	14,92%	14,90%	14,70%	14,57%	14,28%
22	15,05%	15,30%	15,35%	15,20%	15,19%	15,08%	14,87%	14,76%	14,54%	14,30%
23	15,05%	15,35%	15,36%	15,29%	15,24%	14,98%	14,81%	14,69%	14,48%	14,50%
24	15,04%	15,43%	15,34%	15,26%	15,12%	15,03%	14,90%	14,64%	14,51%	14,35%
25	15,05%	15,33%	15,28%	15,23%	15,09%	15,00%	14,88%	14,66%	14,53%	14,40%
26	15,03%	15,39%	15,38%	15,24%	15,28%	14,96%	14,87%	14,60%	14,44%	14,35%
27	15,02%	15,25%	15,38%	15,18%	15,21%	14,95%	14,86%	14,72%	14,53%	14,32%
28	15,02%	15,38%	15,44%	15,22%	15,26%	14,94%	14,85%	14,69%	14,51%	14,25%
29	15,01%	15,33%	15,38%	15,29%	15,14%	15,04%	14,89%	14,72%	14,49%	14,40%
30	14,96%	15,31%	15,27%	15,25%	15,20%	15,12%	14,89%	14,72%	14,48%	14,32%
31	14,94%	15,40%	15,28%	15,27%	15,07%	15,01%	14,91%	14,81%	14,54%	14,32%
32	14,97%	15,34%	15,32%	15,22%	15,15%	15,04%	14,82%	14,75%	14,46%	14,45%
33	14,96%	15,28%	15,44%	15,30%	15,16%	14,95%	14,86%	14,74%	14,46%	14,44%
34	14,93%	15,29%	15,36%	15,28%	15,30%	14,91%	14,84%	14,78%	14,59%	14,41%
35	14,95%	15,24%	15,37%	15,30%	15,11%	14,97%	14,79%	14,80%	14,60%	14,37%
36	14,86%	15,30%	15,31%	15,28%	15,22%	15,09%	14,91%	14,73%	14,49%	14,43%
37	14,86%	15,27%	15,34%	15,35%	15,13%	15,09%	14,83%	14,67%	14,69%	14,46%
38	14,82%	15,32%	15,35%	15,23%	15,26%	15,10%	15,06%	14,75%	14,66%	14,41%
39	14,82%	15,35%	15,46%	15,31%	15,30%	14,99%	14,90%	14,80%	14,62%	14,50%
40	14,79%	15,30%	15,38%	15,29%	15,18%	15,06%	14,93%	14,80%	14,59%	14,56%
41	14,75%	15,28%	15,40%	15,35%	15,15%	15,15%	14,96%	14,92%	14,65%	14,38%
42	14,84%	15,39%	15,41%	15,43%	15,14%	15,15%	15,02%	14,78%	14,53%	14,43%
43	14,83%	15,34%	15,36%	15,20%	15,23%	15,03%	15,07%	14,76%	14,73%	14,47%
44	14,78%	15,27%	15,37%	15,24%	15,29%	15,23%	15,01%	14,78%	14,70%	14,55%
45	14,77%	15,36%	15,33%	15,41%	15,19%	15,17%	15,00%	14,82%	14,68%	14,53%
46	14,70%	15,35%	15,38%	15,23%	15,36%	15,13%	15,06%	14,81%	14,63%	14,62%
47	14,79%	15,28%	15,40%	15,28%	15,23%	15,08%	14,90%	14,82%	14,80%	14,62%
48	14,76%	15,25%	15,32%	15,27%	15,28%	15,12%	14,96%	14,71%	14,75%	14,61%
49	14,68%	15,30%	15,47%	15,45%	15,20%	15,15%	15,00%	14,88%	14,75%	14,61%
50	14,68%	15,33%	15,37%	15,38%	15,21%	15,09%	15,00%	14,96%	14,66%	14,61%

Tabla 10- Búsqueda de grilla para alfa y factores. Fuente: Elaboración propia

En la tabla 10 se tiene los resultados de la grilla para un alfa que varía entre 1 y 10 con el número de factores entre 1 y 50. De los resultados, se ve que los mejores están en el intervalo [2,4] para alfa, mientras que para los factores no hay cambios tan evidentes, pero es destacable que las iteraciones con 1 y 2 factores tienen resultados inferiores a todas las demás realizadas. Dada la grilla, es que para esta y las siguientes iteraciones al modelo, el alfa y número de factores a utilizar quedan definidos por el máximo en R@5 que corresponde a:

Alfa	3
Factores	13
R@5	15,7%

Tabla 11-Máximo búsqueda de grilla. Fuente: Elaboración propia

5.3.1.2. Análisis de resultados y escalabilidad del modelo

K	Aciertos	Aciertos@k	Recall@k	Aciertos clientes @k	Recall clientes @k
1	2.686	2.686	5%	2.686	10%
2	2.084	4.770	8%	4.329	16%
3	1.770	6.540	11%	5.604	21%
4	1.463	8.003	13%	6.570	24%
5	1.348	9.351	16%	7.374	27%

Tabla 12- Análisis resultados factorización de matrices base. Fuente: Elaboración propia

De la tabla de resultados en primer lugar es posible notar que, de acuerdo a la intuición, a medida que el ranking de la recomendación desciende, también lo hace el número de aciertos y por ende el aporte que se genera en el R@K y el RC@K al ir agregando recomendaciones de más abajo de la lista para cada cliente. En el gráfico 5 es posible ver este fenómeno de manera visual, donde ambas curvas de aciertos van disminuyendo su crecimiento al agregar otro producto.

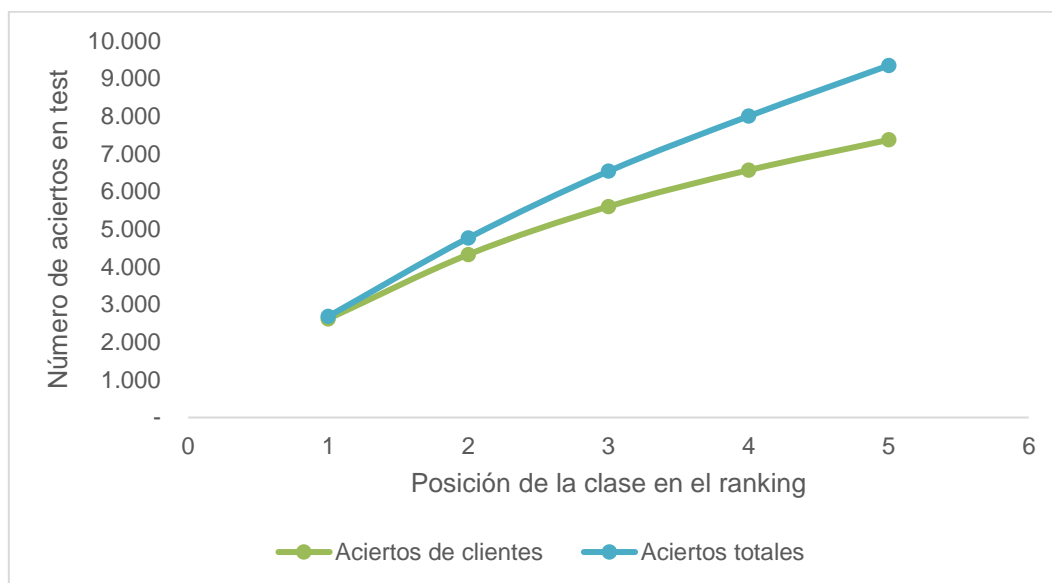


Gráfico 5- Aciertos en test. Fuente: Elaboración propia

Con respecto a la capacidad predictiva del modelo, se tiene que en base al R@5 entrega mejores resultados que el modelo de lo más vendido, específicamente tiene un 28.5% de mejora.

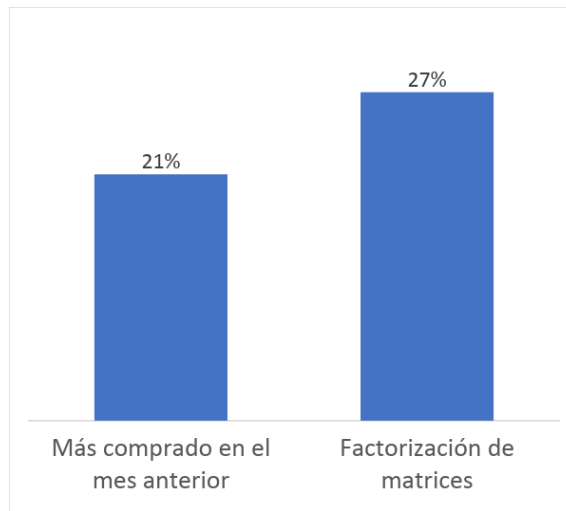


Gráfico 6- Más comprado mes anterior y factorización de matarices. Fuente: Elaboración propia

En cuanto al tiempo de ejecución, el código demora en torno a los 107 segundos aproximadamente en entregar las propuestas de recomendación y los resultados del modelo. Una prueba relevante a realizar corresponde a verificar la escalabilidad del modelo ya que en caso de que sea implementado por la empresa este debe ser lo suficientemente escalable para entregar resultados a un grupo lo suficientemente grande de clientes. Para analizar esta hipótesis se generaron seis muestras del género femenino que varían entre 24.000 y 247.000 clientes y que cumplen con las mismas condiciones que la muestra utilizada para evaluar los modelos.

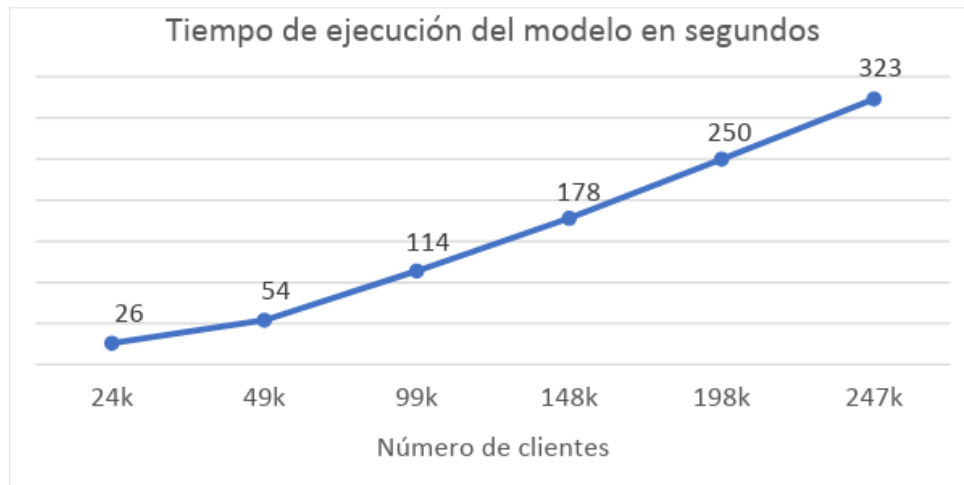


Gráfico 7- Escalabilidad FM. Fuente: Elaboración propia

Del gráfico es posible validar la hipótesis y notar que el modelo escala de manera casi lineal con una tasa de 1.3 segundos por cada cliente nuevo en la base. Por lo tanto, se tiene que el modelo de factorización de matrices es adecuado para un alto volumen de clientes.

5.3.2. Iteración 1: No recomendar clases-marcas ya compradas

Para esta iteración se hace un cambio sencillo a la implementación del modelo anterior, restringiendo las recomendaciones solo a categorías que el cliente no haya comprado. La implementación del script en Python tiene cuatro etapas:

1. Obtención de clase-marca ya comprados por persona para generar una lista con los productos a filtrar para cada cliente.
2. Ejecución del modelo base de factorización de matrices (5.3.1) para obtener 30 recomendaciones ordenadas por cliente.
3. Para cada cliente se eliminan las clase-marca en la lista de filtro.
4. Re-ranking de las clase-marca sobrevivientes en base al ranking original para cada cliente.

5.3.2.1. Análisis de resultados

K	Aciertos	Aciertos@k	Recall@k	Aciertos clientes @k	Recall clientes @k
1	1104	1104	2%	1129	4%
2	943	2047	3%	2017	8%
3	746	2793	5%	2693	10%
4	751	3544	6%	3345	12%
5	631	4174	7%	3883	14%

Tabla 13- Resultados FM sin recomendar clase-marca ya compradas. Fuente: Elaboración propia

De los resultados en la tabla 13 se puede inferir que el modelo tiene un comportamiento similar al anterior, donde a medida que baja la posición en el ranking el número de aciertos desciende. Por otro la eficiencia es inferior al modelo base, teniendo este último un 128% y 92% de mejora en R@5 y RC@5 respectivamente, más aún la iteración sin recomendar productos ya comprados anteriormente entrega peores resultados que el modelo de lo más vendido, el cual tiene un 12% y 21% en cuanto a R@5 y RC@5. Se plantea que estos resultados son acordes con la intuición ya que durante el año un cliente es probable que compre varias veces dentro de las mismas categorías de vestuario y calzado por preferencia personal, caso distinto sería en la recomendación de electrónica o línea blanca, donde se espera que la frecuencia de compra de un producto dentro de esta clase sea baja y por lo tanto recomendar algo comprado hace pocos meses puede no ser una buena estrategia.

	Iteración 1	Modelo base	Lift sobre iteración 1
R@5	7%	16%	128%
RC@5	14%	27%	92%

Tabla 14- Lift iteración 1. Fuente: Elaboración propia

Finalmente, sobre no recomendar productos que el cliente haya comprado anteriormente, se concluye que afecta negativamente la efectividad de las recomendaciones de vestuario y calzado, por lo tanto, en las siguientes iteraciones no habrá cambios sobre el modelo base y será posible recomendar productos ya comprados anteriormente por el cliente.

5.3.3. Iteración 2: Incluir la estacionalidad de las clases

5.3.3.1. Desarrollo de metodología

La incorporación de estacionalidad en las recomendaciones se realiza después de la obtención de recomendaciones desde el modelo base, utilizando el indicador de proporción con respecto a la media anual visto anteriormente (sección 5.1.2). Para realizar el ajuste se obtienen 100 recomendaciones para cada cliente, que además de tener el ranking se conserva el score utilizado para generarlo. Este score se pondera directamente con el indicador de proporción a nivel de cada clase (no depende de la marca), el cual es mayor a 1 cuando las ventas de un mes en particular son mayores al promedio anual y menor a uno en caso contrario. Con el nuevo puntaje ponderado se procede a generar un nuevo ranking donde se seleccionan las 5 primeras clase-marca como recomendaciones para cada cliente.

Como ejemplo del efecto de las ponderaciones, en las ilustraciones 1 y 2 se ven dos sets de recomendaciones distintas para el mismo cliente en el mes de noviembre 2018. Una utilizando el ponderador de estacionalidad y otra no, en estas es posible ver como los sweaters y algunas clase-marca de poleras ya no son recomendados en el modelo con estacionalidad, en su lugar son reemplazados por poleras deportivas y camisas.



Ilustración 1- Factorización con estacionalidad

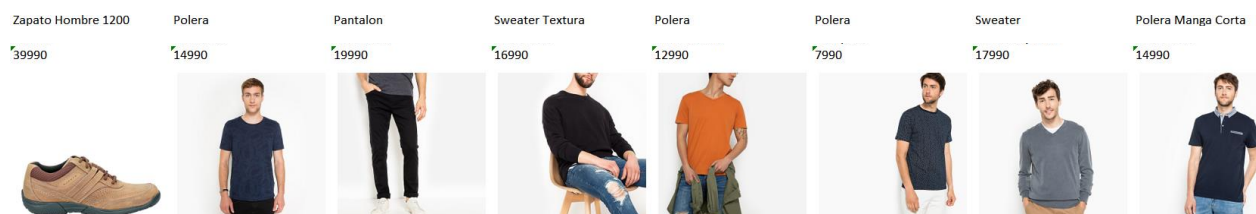


Ilustración 2- Factorización sin estacionalidad

5.3.3.2. Análisis de resultados

K	Aciertos	Aciertos@k	Recall@k	Aciertos clientes @k	Recall clientes @k
1	2886	2886	5%	2934	11%
2	2198	5084	9%	4801	18%
3	1694	6777	11%	6071	22%
4	1483	8260	14%	7084	26%
5	1332	9592	16%	7917	29%

Tabla 15- Resultados factorización de matrices con estacionalidad. Fuente: Elaboración propia

De los resultados en la tabla 15 obtenidos en un tiempo en torno a los 446 segundos, se ve que a medida que baja la posición en el ranking el número de aciertos desciende, por lo tanto, el orden en que se seleccionan los productos para cada cliente tiene sentido. Con respecto a la eficacia se tiene que para el R@5 a pesar de que es tiene levemente más aciertos no alcanza a ser más de un punto porcentual sobre el modelo base, por otro lado, el RC@5 entrega un mejor resultado que el modelo base de factorización de matrices, específicamente tiene una lift de un 7%. Esta mejora es esperable ya que de acuerdo a lo visto en la sección 5.1.2 existen categorías de productos que tienen una variación considerable en sus ventas durante el año, donde por ejemplo recomendar productos como un traje de baño en invierno no entregaría buenos resultados.

	Modelo base	Iteración 2	Lift sobre modelo base
R@5	16%	16%	-
RC@5	27%	29%	7%

Tabla 16- Lift estacionalidad. Fuente: Elaboración propia

Para corroborar que el RC@5 del modelo con estacionalidad es significativamente distinto que el modelo base, se realiza un test de proporciones de dos colas, donde la hipótesis nula propone que las proporciones son iguales y la alternativa que son distintas. En la tabla 17 se pueden ver los resultados de este test, donde con un p-valor menor a 0.01 se rechaza la hipótesis nula y se permite señalar que las proporciones son distintas al 99% de confianza.

Proporción 1	Proporción 2	N1	N2	Z calculado	P-valor
29%	27%	26860	26860	5.1621	<0.00001

Tabla 17- Test proporciones modeo con estacionalidad. Fuente: Elaboración propia

Finalmente, sobre el modelo con estacionalidad agregada a las recomendaciones, se concluye que la inclusión de estacionalidad aporta mejoras al modelo base en cuanto al RC@5, por lo que en la siguiente iteración será considerada al recomendar productos de vestuario y calzado.

5.3.4. Iteración 3: Añadir gasto pasado del cliente

5.3.4.1. Desarrollo de la metodología

La incorporación del gasto pasado del cliente tiene 3 pasos:

1. Segmentación de las marcas de acuerdo a su precio promedio de venta para cada clase.
2. Determinación de la proporción de las compras pasadas de cada cliente en los segmentos del punto 1.
3. Uso de la proporción para ponderar el puntaje de las recomendaciones de factorización de matrices para cada cliente.

La segmentación de las marcas se realizó en Python mediante el algoritmo k-means y el uso de cuatro segmentos, agregar un segmento muy alto es relevante debido a que algunas marcas de vestuario exclusivas como Guess tienen una gran diferencia de precio con respecto a las marcas seguidoras, lo que en una segmentación k-means con 3 grupos (bajo-medio-alto) provoca que se genere un segmento alto solo con estas marcas. En anexo B se tiene un ejemplo de esta diferencia de precios con la clase jeans de la sublínea “Bottoms señora”, donde Guess que está en el grupo muy alto, tiene un precio promedio de 119.000 CLP mientras que la marca siguiente dentro de la clase tiene un precio de 41.900 CLP. Otro punto relevante es que la segmentación se hace dentro de cada clase y no se realiza solo en base al precio promedio de la marca, esta decisión se debe a que algunas marcas tienen distintos estatus dependiendo de la clase.

Grupo	Segmento de precio
0	Bajo
1	Medio
2	Alto
3	Muy alto

Tabla 18- Nombres segmentos. Fuente: Elaboración propia

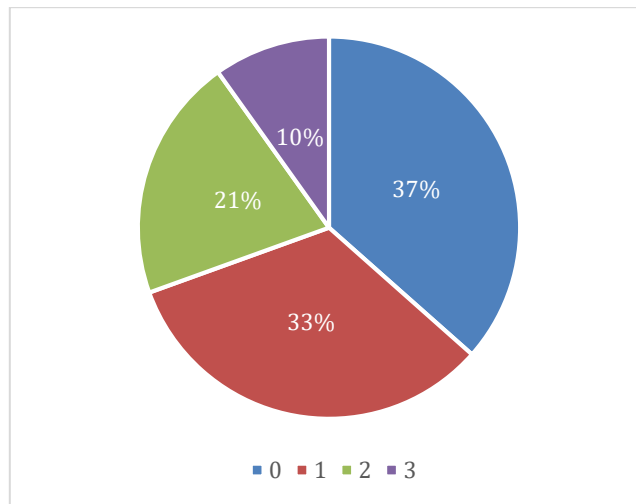


Gráfico 8- Participación de cada segmento en el número de clases-marca totales

Con las clases-marca segmentadas, se etiquetan las compras en el periodo de entrenamiento de cada cliente de la muestra en base al segmento que pertenecen. Luego la proporción de compras en cada segmento por cliente se calcula:

$$\text{Proporción de compras}_s = \frac{\# \text{ Compras en segmento}_s}{\# \text{ Compras totales del cliente}}$$

De este modo para cada cliente se tiene un porcentaje de compras en cada segmento, así, si un cliente tiene 0% de gasto en alguno de los segmentos significa que hacerle recomendaciones de productos en él probablemente entregue malos resultados. En la tabla 19 se puede ver el promedio de la proporción de compras de los clientes para cada segmento, de él se desprende que en general el gasto de los clientes está dentro de los segmentos bajo y medio, con una concentración de un 88% mientras que en los segmentos altos y muy alto el volumen de transacciones es menor.

Segmento	Promedio
Bajo	53%
Medio	35%
Alto	10%
Muy Alto	2%

Tabla 19- Promedio de share de los clientes de la muestra en cada segmento. Fuente: Elaboración propia

El último paso a realizar para la implementación corresponde a ponderar la proporción de compras por el puntaje obtenido después de ejecutar el modelo con estacionalidad (recordando que en 5.3.3 se concluye que el modelo con estacionalidad entrega mejores resultados que el modelo base), así este puntaje se ajustará al gasto del cliente en cada

uno de los segmentos de manera proporcional a su gasto en ellos durante el periodo de entrenamiento.

5.3.4.2. Análisis de resultados

K	Aciertos	Aciertos@k	Recall@k	Aciertos clientes @k	Recall clientes @k
1	2606	2450	4%	2649	10%
2	1813	4419	7%	4197	16%
3	1507	5926	10%	5319	20%
4	1304	7230	12%	6207	23%
5	1114	8344	14%	6903	26%

Tabla 20- Resultados iteración 3, incluir gasto del cliente. Fuente: Elaboración propia

Con un tiempo de ejecución en torno a los 1.070 segundos se obtienen los resultados de la tabla 20, en ella se puede ver que nuevamente se cumple la tendencia de disminuir el número de aciertos a medida que el ranking del producto empeora. En comparación con iteraciones anteriores se tiene que agregar el share de gasto del cliente entrega peores resultados que el modelo base y por ende que el modelo con estacionalidad, donde este último tiene una mejora de un 14% y 11% en cuanto a R@5 y RC@5 respectivamente.

	Iteración 3	Iteración 2	Lift de iteración 2 sobre iteración 3
R@5	14%	16%	14%
RC@5	26%	29%	11%

Tabla 21- Lift iteración 3. Fuente: Elaboración propia

En base a esto, sobre el utilizar el gasto pasado del cliente, se tiene que empeora los resultados con respecto a utilizar solo el modelo con estacionalidad, por lo tanto, este último sigue siendo la mejor alternativa dentro de factorización de matrices.

Por último, se cree que algunas de las causas que provocan el empeoramiento del modelo pueden ser:

- El modelo de factorización de matrices captura correctamente las preferencias de los clientes de acuerdo a el precio de los productos de su gasto pasado y utilizar esta información de forma posterior solo agrega ruido.
- Que la segmentación por precios dentro de las clases no esté bien abordada. En tal caso se requiere desarrollar un trabajo enfocado en este punto para lograr una buena implementación.
- La segmentación de precios genera una disminución en las recomendaciones de marcas masivas, impactando en el resultado final.

5.3.5. Conclusiones factorización de matrices

En el gráfico 9 se tienen los resultados del R@5 para los modelos de factorización de matrices y el modelo de benchmark. En él se puede ver directamente que el mejor modelo corresponde al de factorización de matrices con estacionalidad en el que se permite recomendar clases-marca ya comprados anteriormente a los clientes. También de las secciones anteriores se tiene que para todos los modelos de factorización de matrices el orden con el que se recomiendan los productos está relacionados con los aciertos que logran.

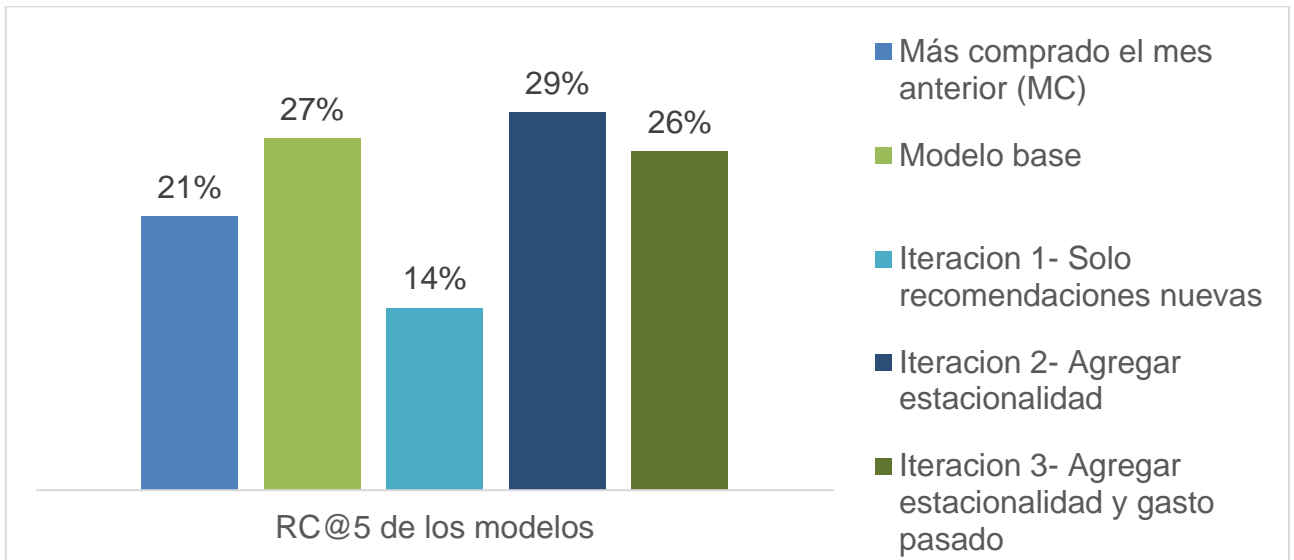


Gráfico 9- Resultados factorización de matrices

Con respecto a la eficiencia de los modelos, en el gráfico 10 se muestra el tiempo en segundos que demoran en generar las recomendaciones para la muestra las distintas variantes. Como observación, estos tiempos no consideran la preparación previa de los datos, por lo que el benchmark del más comprado tiene 0 segundos dado que consiste solo en una consulta liviana en SQL. También el modelo base y la iteración 1 tienen tiempos de ejecución en torno a números similares, debido a que el filtro se realiza de forma instantánea en la muestra. Los mayores tiempos ocurren para las iteraciones 2 y 3, debido a que posterior a la ejecución del modelo base la iteración número 2 realiza el procedimiento de añadir estacionalidad y la iteración 3 debe añadir estacionalidad a los resultados y además incorporar el gasto pasado del cliente. Por último, con respecto a la escalabilidad del modelo en base al número de clientes utilizados, en la sección 5.3.1.2 se comprueba que el crecimiento en tiempo del modelo base tiene un comportamiento similar a un crecimiento lineal, el cual debido al modo de implementación de las iteraciones debería ser heredado por estas (el comportamiento lineal, no la tasa de crecimiento).

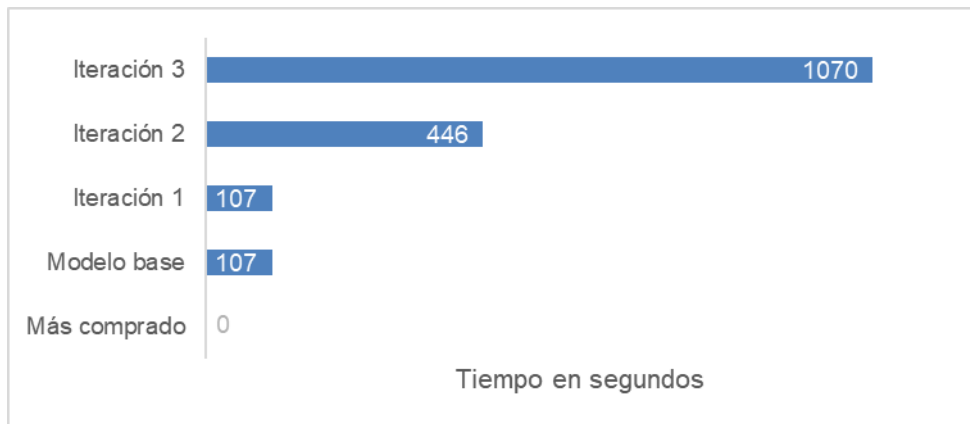


Gráfico 10- Tiempo de ejecución de los modelos. Fuente: Elaboración propia

Como conclusión sobre factorización de matrices, se tiene que dados:

1. Los resultados en cuanto a R@5.
2. El tiempo de ejecución razonable y la escalabilidad de este al aumentar el número de clientes.

Se selecciona el modelo de factorización de matrices con estacionalidad y que permite recomendar productos ya comprados como la mejor de las implementaciones dentro de factorización de matrices. Además, para futuras implementaciones de estas metodologías se sugiere:

- Analizar el impacto que genera en la eficacia agregar información sobre visitas en la página web.
- Revisar la implementación del gasto pasado y comprobar otras formas de incluirlo en el modelo. Con esto se puede verificar si la baja en la efectividad se debe a que el modelo base captura de buena forma el atributo precio o si la iteración probada en este trabajo solo agrega ruido.
- Probar el uso de un alfa distinto para clase-marca masivas para disminuir el gran peso que tienen en las recomendaciones.

5.4. Recomendaciones de KNN

5.4.1. Modelo base

El modelo KNN desarrollado corresponde a una implementación en SQL, específicamente se ejecuta en el entorno de big data conformado por cloudera hive y cloudera impala para tener más velocidad en las consultas.

Y como se menciona en la metodología, para encontrar los vecinos cercanos de cada cliente se utiliza la distancia de Jaccard:

$$\begin{aligned} \text{Distancia Jaccard} &= 1 - \text{similitud Jaccard} \\ &= 1 - \frac{\text{Cantidad de clase marca iguales comprados por ambos (intersección)}}{\text{Cantidad de clase marca comprados por ambos (unión)}} \end{aligned}$$

Para realizar las recomendaciones se consideran como vecinos cercanos los 500 con menor distancia Jaccard y posteriormente se realiza la recomendación en base a las clases-marca más compradas por los vecinos en el mes anterior, cabe señalar que esta aproximación para recomendar los productos ya tiene considerado estacionalidad debido a que las recomendaciones son obtenidas desde compras realizadas recientemente por los clientes.

5.4.1.1. Análisis de resultados

Con un tiempo de ejecución en torno a los 70 minutos y un total de 2.849.936.253 distancias calculadas entre mujeres y 300.014.760 entre hombres para los 99.994 clientes, se obtienen los resultados de la tabla 22. Como observación se tiene que para el caso de los modelos basados en KNN, no es posible ver los resultados de los aciertos para menos de 5 recomendaciones debido a la forma en que se programó inicialmente el código para medir el desempeño.

	Aciertos @5	R@5	Aciertos clientes @5	RC@5
Benchmark - más vendido	6866	12%	5678	21%
KNN	10.372	17%	8.333	31%

Tabla 22-Resultados KNN más vendido mes anterior. Fuente: Elaboración propia

De los resultados de esta implementación se puede ver de forma directa que el modelo entrega mejores predicciones que el Benchmark, específicamente tiene un 42% y un 48% de lift en cuanto a R@5 y RC@5 respectivamente. Además, tiene un desempeño un 8% mejor que el de factorización de matrices con estacionalidad. Como punto negativo la ejecución del modelo tiene un aumento de tiempo de 39 veces aprox. con respecto al modelo base de factorización de matrices, donde el principal costo en tiempo corresponde al número de distancias que se necesitan calcular, donde para N clientes corresponde a:

$$\text{Combinatoria de pares con } N \text{ elementos} = \frac{N!}{(N-2)! * 2!} = \frac{1}{2} * N * (N-1)$$

De acuerdo a estos cálculos el número de distancias a calcular para el problema no solo es grande, sino que su tasa de crecimiento no es lineal. Para un N lo suficientemente grande, se puede asumir que $N \cong N - 1$, así el orden de complejidad del número de distancias a calcular es $O(N^2)$ en notación de Landau. En el gráfico 11 se puede ver este crecimiento del número de distancias a calcular a medida que crece el número de clientes.

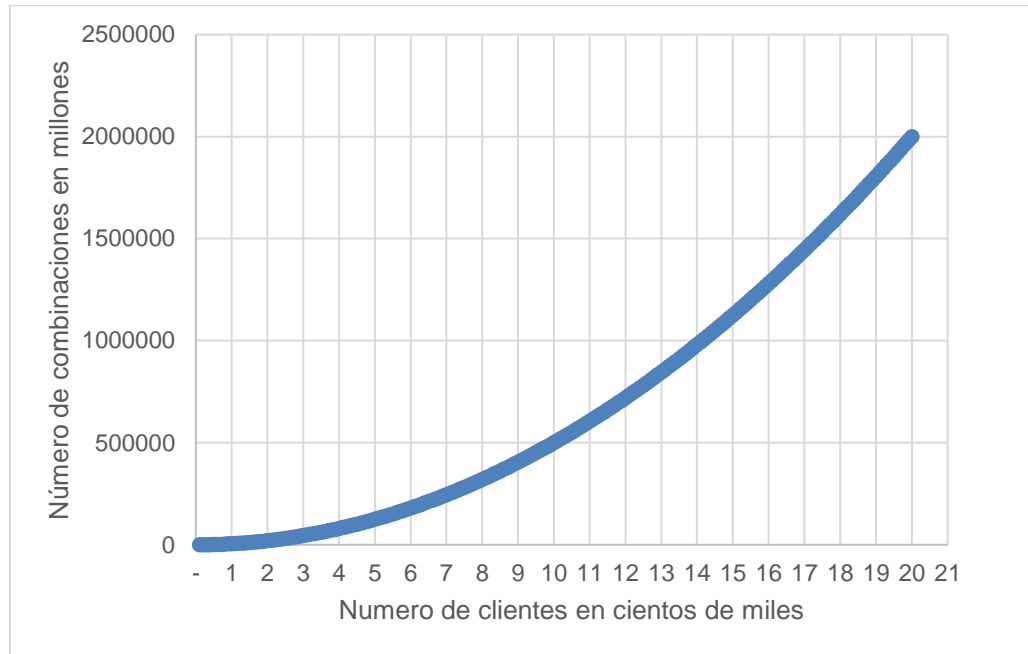


Gráfico 11- Crecimiento de combinaciones KNN. Fuente: Elaboración propia

Entonces sobre el modelo base de KNN, se concluye que tiene mejores resultados que el mejor modelo de factorización de matrices a costa de aumentar el tiempo de ejecución y con la diferencia de que a medida que se aumente el número de clientes a recomendar el tiempo de factorización de matrices crece de forma constante mientras que el de KNN “explota”.

En las siguientes secciones, se presentan tres variantes de KNN, donde se busca mejorar la eficacia del modelo.

5.4.2. Iteración 1: Recomendar en base a lo que compran los vecinos por sobre las compras generales

Esta modificación se realiza luego de calcular las distancias entre clientes. La variación consiste en que en lugar de recomendar lo más comprado por la vecindad, se recomienda en base a cuanto de una clase-marca ha comprado la vecindad por sobre las compras

promedio de todos los clientes, de este modo se busca darles más importancia a los gustos específicos del conjunto de vecinos. A continuación, se ven las fórmulas utilizadas para los cálculos:

$$\text{Promedio clasemarca} = \frac{\text{Ventas totales de la clasemarca en mes anterior}}{\text{Numero de clientes que compraron algo el mes anterior}}$$

$$\text{Promedio vecindad clasemarca} = \frac{\text{Ventas de la clasemarca en la vecindad el mes anterior}}{\text{Número de clientes de la vecindad que compraron algo el mes anterior}}$$

$$\text{Mas comprado de la vecindad sobre el promedio} = \frac{\text{Promedio vecindad clasemarca}}{\text{Promedio clasemarca}}$$

Con un tiempo de ejecución en torno a los 76 minutos, se obtienen los resultados de la tabla 23. En base al R@5 y RC@5, es directo ver que la iteración tiene un peor desempeño que el modelo base de KNN, más aún, entrega peor resultados que el benchmark de lo más vendido para el mismo género el mes anterior. Estos resultados se cree que ocurren debido a que al intentar obtener recomendaciones “de nicho” para los clientes de acuerdo a su vecindad, se pierden los productos masivos que aportan gran parte de la venta mensual, como es posible ver en el modelo de benchmark que genera un 21% RC@5 solo recomendando en base a estos productos.

	Aciertos @5	R@5	Aciertos clientes @5	RC@5
Benchmark - más vendido	6866	12%	5678	21%
KNN base	10.372	17%	8.333	31%
KNN vecindad/promedio	5196	9%	4548	17%

Tabla 23- KNN lift. Fuente: Elaboración propia

5.4.3. Iteración 2: No considerar transacciones con ofertas en la selección de la clase-marca

Del mismo modo que la prueba anterior, en esta se realiza un cambio en la forma de seleccionar los productos dada la vecindad. La implementación en este caso corresponde a no considerar transacciones que tengan un precio de venta menor a un 50% del precio normal al calcular las clase-marca más compradas por los vecinos, así se busca limpiar compras guiadas por las ofertas en lugar de la preferencia por alguna clase-marca en particular. En la tabla 24 se tienen los resultados obtenidos en un tiempo de ejecución similar al modelo base, donde se puede ver que hay una leve mejora en la eficacia sobre este.

	Aciertos @5	R@5	Aciertos clientes @5	RC@5
Benchmark - más vendido	6866	12%	5678	21%
KNN base	10.372	17%	8.333	31%
KNN sin ofertas >50%	10.441	18%	8.456	32%

Tabla 24- KNN sin ofertas 50%. Fuente: Elaboración propia

Luego, para corroborar si realmente la diferencia de RC@5 entre los modelos es significativa, se hace un test z de proporciones de dos colas. De los resultados disponibles en la tabla 25, se tiene que con un P-valor menor a 0,05 se rechaza la hipótesis nula de igualdad de proporciones al 95% de confianza, por ende, se concluye que el modelo KNN que no considera ofertas mayores al 50% para analizar los productos más comprados por la vecindad entrega mejores resultados que el modelo KNN base.

Proporción 1	Proporción 2	N1	N2	Z calculado	P-valor
31%	32%	26860	26860	2,494	0,12

Tabla 25- Test de proporciones KNN sin ofertas 50%. Fuente: Elaboración propia

5.4.4. Conclusiones KNN

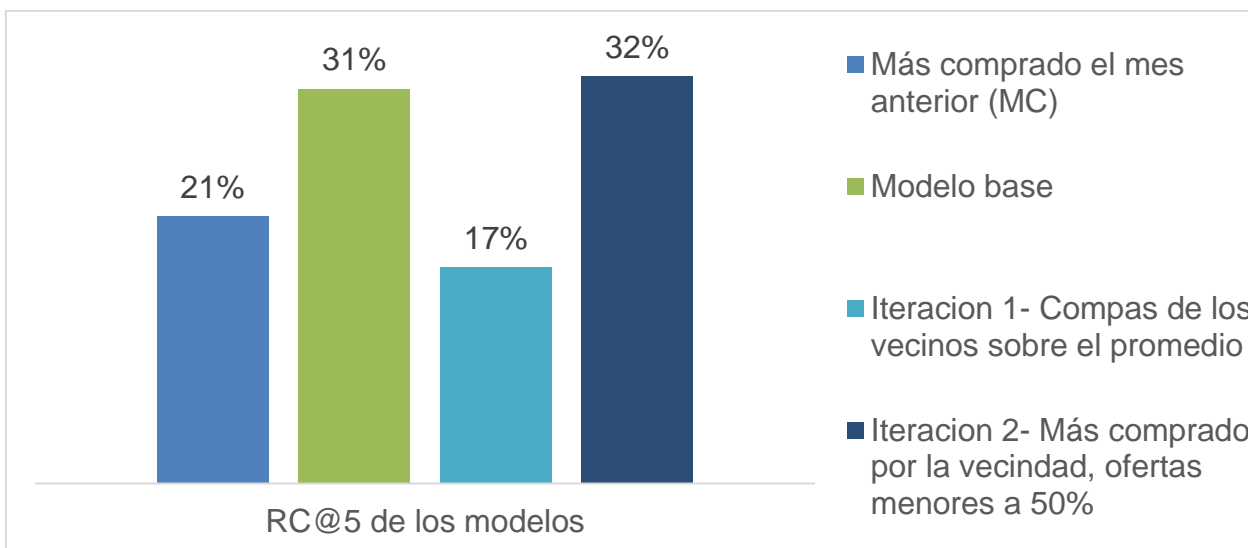


Gráfico 12-Resultados KNN. Fuente: Elaboración propia

En el gráfico 12 se tienen los resultados del RC@5 para los modelos de KNN y el modelo de benchmark. En él se puede ver directamente que el mejor modelo corresponde al KNN

con distancia Jaccard que no utiliza transacciones de productos vendidos a menos de un 50% del precio normal. Con respecto a su eficiencia, se concluye que todos los modelos de KNN implementados en SQL mediante fuerza bruta, tienen un tiempo de ejecución mayor a 70 minutos, donde este no escala de manera lineal al aumentar el número de clientes, lo que podría generar un problema al querer realizar una implementación de KNN a un gran número de clientes del negocio.

5.5. Selección de modelo a utilizar en fase experimental

Con todos los modelos implementados y los resultados obtenidos, es necesario elegir uno para el desarrollo de la sección experimental, específicamente se comparan los mejores de cada familia de filtros colaborativos analizados. De la tabla 26 y el gráfico 13 se puede concluir que el mejor modelo en base al RC@5 corresponde al proveniente de KNN, que tiene una mejora de un 10% sobre factorización de matrices y un 52% sobre el más comprado el mes anterior. Por contraparte, como se explicitó en la sección anterior, el tiempo de ejecución de un modelo KNN es mucho mayor que de factorización de matrices y además no tiene la propiedad de escalar con el aumento del número de clientes como si lo tienen las otras dos opciones. Entonces, en base al defecto de los modelos de KNN que conllevaría un alto costo para ser implementado a gran escala en el negocio y a que la mejora sobre factorización de matrices no es tan sustancial, se concluye que el modelo seleccionado para ser probado experimentalmente y con ello medir el aporte que podría generar en el negocio, corresponde al de factorización de matrices con estacionalidad.

	Modelo de Benchmark más comprado	Factorización de matrices con estacionalidad	KNN sin ofertas mayores a un 50%
R@5	12%	16%	18%
RC@5	21%	29%	32%
Tiempo ejecución	-	< 7 mins	70 mins aprox
Escalable	Si	Si	No

Tabla 26- Mejor modelo. Fuente: Elaboración propia

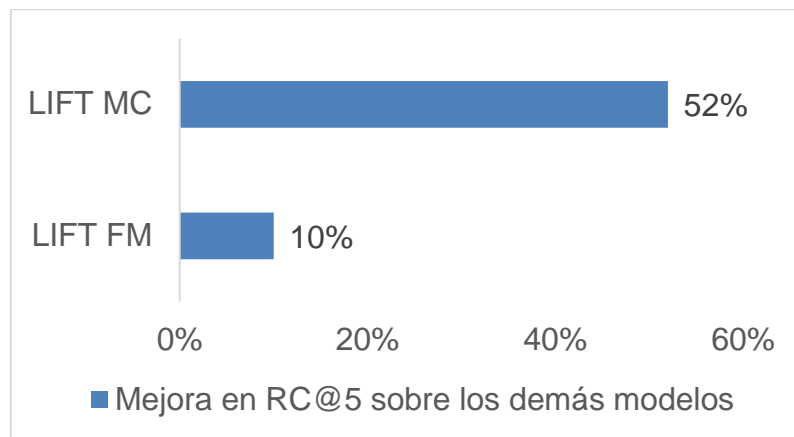


Gráfico 13- Lift KNN sobre el modelo de lo más comprado (MC) y factorización de matrices (FM). Fuente: Elaboración propia

6. Desarrollo metodología fase de experimentación

6.1. Preparación del envío

6.1.1. Diseño de la gráfica

Como se menciona en secciones anteriores, el desarrollo del experimento se realiza en conjunto con el área de inteligencia web. Con ellos se acuerda el formato y diseño que tendrá el correo para los envíos, siguiendo las restricciones y características definidas a continuación, estas son del tipo restricción de la empresa (DE) o propuesta para el trabajo (PT):

- En el correo se deben incluir recomendaciones de calzado de manera obligatoria, principalmente porque esta línea es de las que tienen mayor relevancia en las ventas por internet. (RE)
- Debido a que un email abierto por un cliente debe tener una coherencia interna y “contar un relato”, se restringen las recomendaciones de vestuario solo a las líneas: hombre, mujer, juvenil hombre y juvenil mujer según corresponda. Se dejan fuera las líneas de deporte y accesorios. (RE)
- Con el mismo objetivo de buscar una coherencia interna, se deben recomendar solo las dos mejores clase-marca de vestuario y de calzado. Esto busca no ampliar el espectro de recomendaciones con algunas con menor nivel de seguridad (peor ranking en factorización de matrices). (RE)
- Para cada clase-marca se muestran 3 SKU distintos y un botón para acceder a todos los productos de la marca en la página del negocio. (PT)
- El precio de los productos no es mostrado en el correo, esto tiene como objetivo evitar tráfico generado por alguna oferta en lugar del interés por los productos. (PT)

Entonces, en base a estas características se tiene el diseño final del template dinámico del correo, un ejemplo con productos seleccionados se puede ver en la ilustración 3,

mientras que las secciones que componen correo y su nivel de personalización están descritas en la tabla 27.

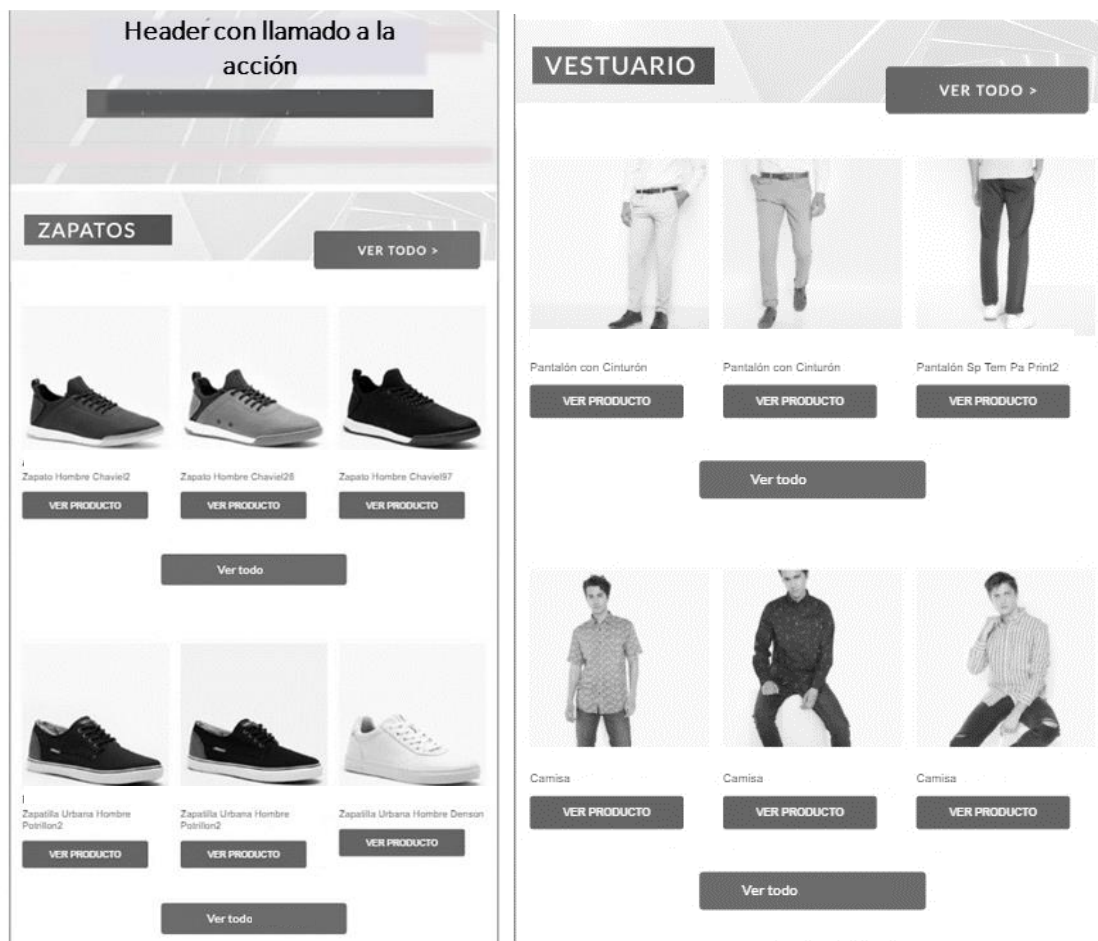


Ilustración 3- Template dinámico del correo. Fuente: Elaboración propia

Sección	Personalización
Header	Estático no personalizable
Banner calzado	Estático no personalizable
Recomendaciones clase-marca 1 calzado	Dinámico calzado
Recomendaciones clase-marca 2 calzado	Dinámico calzado
Banner vestuario	Estático no personalizable
Recomendaciones clase-marca 1 vestuario	Dinámico vestuario
Recomendaciones clase-marca 2 vestuario	Dinámico vestuario
Footer	Estático no personalizable

Tabla 27- Secciones que componen el correo. Fuente: Elaboración propia

6.1.2. Elección de clientes

Para la elección de los clientes con quienes se realizará el experimento se fijan una serie de condiciones que corresponden a:

- Tener asociado un email de contacto, que debe tener al menos una apertura detectada durante el 2019.
- Deben tener al menos 5 compras en clases-marcas distintas de vestuario y calzado entre marzo 2018 y febrero 2019.
- Un openrate mínimo de un 10% para asegurar aperturas.

En base a estas restricciones, se selecciona aleatoriamente un conjunto de 200.000 clientes para ser divididos en grupo de control y tratamiento. Esta división se realiza mediante la técnica de resampleo aleatorio implementada en R con el paquete “mindiff” utilizando como input para la división de grupos las siguientes variables:

- Cantidad de compras por cliente en el periodo de entrenamiento del modelo
- Gasto de los clientes en vestuario y calzado
- Openrate de correos enviados
- Tasa de clicks en correos enviados
- Cantidad de hombres y mujeres en cada grupo

Con estas variables se busca separar los clientes en dos grupos de 100.000 cada uno, de modo que ambos tengan una media y desviación estándar similares en las variables continuas, mientras que para cantidad de hombres y mujeres tengan no más de 100 clientes de diferencia entre grupos. Entonces, con 200 iteraciones realizadas y con la mejor asignación encontrada en el número 194, se tiene los resultados de los grupos en la tabla 28, en este caso las variables cantidad de compras promedio, gasto promedio, openrate promedio y ctr promedio, están transformadas en base una división por el valor del grupo de control por confidencialidad, aun así, es posible apreciar la similitud entre los grupos.

Grupo	Cientes	Cantidad de compras promedio	Gasto promedio	Openrate promedio	Ctr promedio	Cantidad de hombres	Cantidad de mujeres
Tratamiento	100.000	1	1.0016	0.997	1.0011	25.649	74.351
Control	100.000	1	1	1	1	25.722	74.278

Tabla 28- Grupo control y tratamiento

Para corroborar que ambos grupos no sean significativamente distintos en las variables utilizadas, se realizan pruebas de dos colas para verificar diferencia de medias y de varianza para las variables continuas. En la tabla 29 se pueden ver los resultados, donde de acuerdo al p-valor es posible afirmar con un 95% de confianza, que no hay suficiente evidencia para rechazar las hipótesis de igualdad de varianza y media entre ambos conjuntos.

Variable	Prueba de Levene			Prueba t de diferencia de medias		
	Estadístico	P-valor	Se rechaza H0 al 95%	Estadístico	P-valor	Se rechaza H0 al 95%
Gasto	0,001	0,98	No	0,369	0,71	No
Openrate	1,435	0,23	No	0,793	0,43	No
Ctr	0,206	0,65	No	0,291	0,77	No
Cantidad de compras	0,041	0,84	No	0,027	0,98	No

Tabla 29- Pruebas de hipótesis igualdad de grupos. Fuente: Elaboración propia

Para la validar la igualdad de proporción de mujeres entre cada grupo, se hace un test z de proporciones para muestras independientes con dos colas, entregando como resultados un estadístico Z de 0,37 y un p-valor de 0,7 con esto se puede afirmar con un 95% de confianza que no hay evidencia estadística para verificar que la proporción de mujeres entre ambos grupos es distinta.

6.1.3. Obtención de recomendaciones

Con los grupos ya divididos, se necesita obtener dos tipos de recomendaciones distintas. Unas seleccionadas por el área comercial y otro resultado del modelo de Factorización de matrices.

Para las recomendaciones del grupo de control, se comunica a los product manager de las líneas de vestuario y calzado sobre el experimento y como se busca medir el impacto que tiene el uso de recomendaciones personalizadas por sobre las comunicaciones masivas que se realizan actualmente. Luego se les solicita que seleccionen 2 clases-marcas para vestuario y 2 para calzado, que creen que venderán bien en marzo 2019. Además, para cada una se solicitan 6 SKU, 3 para mostrar en el correo y 3 de reserva en caso de que alguno sea despublicado de la página por falta de stock. En la tabla 30 se pueden ver las clase-marca recomendadas por el área comercial, las cuales pertenecen principalmente a marcas para el público masivo.

	Masculino	Femenino
Clase-marca calzado	Zapato-MPH (1)	Zapato-MPM (1)
Clase-marca calzado	Zapatilla-MPH (2)	Zapato-MPM (2)
Clase-marca vestuario	Pantalón-MPH (3)	Chaqueta-MPM (2)
Clase-marca vestuario	Camisa-MPH (4)	Denim-MPM (3)

Tabla 30- Recomendaciones control, MPM y MPH representan marcas exclusivas de hombre y mujer, la numeración es válida solo para esta sección. Fuente: Elaboración propia

Para las recomendaciones personalizadas primero se deben definir las clase-marca disponibles para recomendar, ya que no todas las que vende el negocio y se utilizan para entrenar el modelo, están disponibles en la página web. Otro punto que afecta el número de clases-marca disponibles, está relacionado con la selección de SKU, ya que se necesita que las seleccionadas tengan disponible un mínimo de 6 distintos publicados en la página web para poder ser seleccionadas para el correo.

La selección de los 6 SKU de cada clase-marca se realiza entre los navegados por clientes en la página tres semanas antes de la preparación del envío (15/02/2019 al 07/03/2019). Esta selección se rige en base a las siguientes condiciones, para no alterar la selección de productos (RS) o deben cumplir exigencias comerciales (EC):

- Que las visitas web estén sobre el percentil 25 de la clase-marca (RS)
- Precio al 08/03/2019 mínimo de 9.990 para vestuario y 19.990 para calzado, exigido por la empresa (RC)
- Con los precios al 08/03/2019, el precio del SKU sea mayor al $\text{Precio promedio de la clase-marca} - 2 * SD \text{ precio clase-marca}$, para no recomendar productos muy baratos dentro de una clase-marca que supere los precios mínimos con facilidad. (RS-RC, debido a que para la empresa no es de interés recomendar productos con ofertas demasiado bajas y por otra parte estas ofertas pueden alterar las ventas de los productos de sobremanera)

Luego, para elegir entre los sku de una clase-marca para el desarrollo de la memoria, se proponen: las visitas web, la tasa de conversión de visitas en órdenes y el precio, como variables de interés para realizar la selección. Las dos primeras representan interés por parte de los clientes y el precio el interés de la empresa. Finalmente, para cada SKU dentro de una clase-marca se calcula el promedio simple de las tres variables normalizadas entre 0 y 1 dentro de su clase-marca, para posteriormente elegir los 6 con mayor promedio.

Con esto la cantidad de clase-marca publicadas en la web al momento de realizar las recomendaciones y que cumplen con la restricción de tener 6 o más SKU que cumplen las condiciones antes señaladas se pueden ver a continuación.

Vestuario mujer	220
Calzado mujer	199
Vestuario hombre	117
Calzado hombre	102

Tabla 31- Clases-marca disponibles. Fuente: Elaboración propia

Con el conjunto de clase-marca disponibles para recomendar ya definido, se ejecuta el modelo de factorización de matrices con estacionalidad utilizando todas las compras de vestuario y calzado (sean o no de clases disponibles para recomendar) de los 200.000 clientes. Luego, dentro de las clase-marca que se encuentran en el conjunto de disponibles para recomendar, se eligen las dos con mejor ranking en calzado y en vestuario para cada cliente del grupo de tratamiento, generando un total de 24.427 combinaciones distintas para los 100.000 clientes, de las cuales 13.023 fueron generadas para un cliente único.

Por último, teniendo los grupos de control y tratamiento para el experimento, y las recomendaciones de cada uno conformadas, mediante el software Exacttarget se programa el envío el día viernes 8 de marzo para el día siguiente a las 9:30 AM.

6.2. Resultados del envío

En esta sección se presentan los resultados en base al CTR, la tasa de conversión y el ARPU. La medición de los resultados del A/B test, se realiza el día viernes 15 de dando así a los clientes 6 días desde el envío de los correos.

Antes de medir los indicadores de desempeño, se hace necesario saber cuántos envíos efectivos ocurren en cada grupo y las tasas de apertura que obtienen los correos. En este caso se espera que en cuanto a openrate entreguen resultados similares, ya que los grupos fueron balanceados para tener una misma tasa de apertura y poseen el mismo asunto.

	Envíos efectivos	Clientes que abren	Openrate
Control	91.836	29.934	32,6%
Tratamiento	91.820	29.521	32,2%

Tabla 32- Resultados aperturas. Fuente: Elaboración propia

En la tabla 32 se tienen los resultados de envíos y aperturas, en este caso se puede notar que ambos grupos tuvieron un 9% de correos sin llegar al destinatario, la principal causa de esta baja corresponde a que la multitienda cuenta con un conjunto de clientes llamados "grupo de control anual" utilizado para medir el desempeño de las campañas

de email marketing y por lo tanto son filtrados automáticamente al realizar los envíos. Con respecto a las aperturas, se puede ver para ambos que la tasa es similar. Para corroborarlo se realiza una prueba z de igualdad de proporciones con dos colas, que entrega un estadístico de 1,83 y un p-valor de 0,067, por lo que con una confianza del 95% no se puede afirmar que hay diferencia estadísticamente significativa entre las proporciones de ambos grupos y por ende se concluye que los grupos se comportaron como se esperaba en cuanto a semejanza de openrate.

Con respecto al CTR, en la tabla 33 se tienen los principales resultados. Donde el grupo de tratamiento obtiene un 34% de lift con respecto al grupo de control, mostrando que las recomendaciones personalizadas efectivamente generaron una diferencia de interés con respecto a las recomendaciones comerciales. Para corroborar esta diferencia se realiza una prueba z de igualdad de proporciones con dos colas, que entrega un estadístico de 21,52 y un p-valor menor a 0.00001, por lo que con una confianza del 95% se puede afirmar que hay diferencia estadísticamente significativa entre las proporciones de ambos grupos. Luego, para verificar que los resultados del grupo de tratamiento son efectivamente mayores, se ejecuta una prueba de 1 cola donde con un estadístico de 21,52 y un p-valor de 0.00001 se puede afirmar a un 99% de confianza que las recomendaciones personalizadas son capaces de entregar mayor tráfico hacia la página web.

	Cientes que abren	Aperturas totales	Clics	CTR
Control	29.934	43.396	7.645	17,6%
Tratamiento	29.521	43.010	10.123	23,5%
Lift	-	-	32%	34%

Tabla 33- CTR. Fuente: Elaboración propia

Otro punto interesante a analizar con respecto a los clics corresponde a como se distribuyen dentro del correo, por lo que usando las definiciones de la tabla 25 se tienen la proporción de clics a continuación.

	Control	Tratamiento
Proporción de clics en: Estático no personalizable	25%	19%
Proporción de clics en: Dinámico vestuario	41%	38%
Proporción de clics en: Dinámico calzado	34%	43%
Clics totales	7.645	10.123

Tabla 34- Distribución de clics. Fuente: Elaboración propia

De la tabla 34, es directo notar que el correo del grupo de control tuvo una mayor proporción de clics en vestuario, mientras que el grupo de tratamiento lo tuvo en calzado. Un análisis más relevante se puede obtener al observar la categoría estático no personalizable y la suma de las categorías dinámicas, aquí se puede observar que en el grupo de tratamiento un 81% de los clics fue en algún producto recomendado mientras que en grupo de control este porcentaje solo es de un 75%. Para verificar que ésta diferencia de clics en las categorías personalizables es significativa, se realiza un test de proporciones con una cola, el cual tiene como hipótesis alternativa que la proporción del grupo de tratamiento es mayor ($81% > 75%$), de este se tienen como resultados un estadístico de 9.6 y un p-valor menor a 0.0001 por lo tanto, hay suficiente evidencia estadística para afirmar con un 99% de confianza, que en el grupo de tratamiento hay una mayor proporción de clics en secciones personalizables del mail.

En base a esto se concluye que el correo con recomendaciones personalizadas no solo tiene más clics que el de recomendaciones comerciales, sino que además tiene una mayor proporción en zonas más relevantes del correo.

Por último, con el software Ominutre, se hace seguimiento a las ventas generadas directamente por un cliente que entró a la web por medio de uno de los correos enviados, en la tabla 35 se presenta el monto vendido, el total de ordenes generadas y los indicadores asociados a este tráfico.

	Aperturas	Ventas	Órdenes	ARPU	Conversión
Control	29.934	\$ 3.187.540	74	\$106	0,25%
Tratamiento	29.521	\$ 3.411.760	123	\$116	0,42%
Lift	-	7%	66%	9%	69%

Tabla 35- Ventas generadas. Fuente: Elaboración propia

De la tabla es directo notar que el correo personalizado genera más ingresos y tiene un lift considerablemente mayor en la tasa de conversión sobre el grupo de control, como contraparte este último tiene un ticket promedio mayor que el grupo de tratamiento (43.074 v/s 27.737). Debido a restricciones al utilizar el programa de seguimiento de las ventas (no es posible calcular la desviación estándar para los test), no es posible obtener la venta individualizada de los clientes, por lo tanto, no se puede realizar un test de diferencia de medias para verificar si la diferencia en el ARPU es estadísticamente significativa. Para la tasa de conversión se realiza un test z de proporciones con dos colas, que da como resultados un estadístico de 3,59 y un p-valor de 0,00034, por lo tanto, es posible rechazar con un 95% de confianza la igualdad de proporciones entre el grupo de control y tratamiento. Luego, para verificar que los resultados de este grupo son efectivamente mayores, se hace un test de proporciones de una cola, el que entrega como resultados un estadístico z de 3.59 y un p-valor de 0.00017, por lo tanto, es correcto señalar con un 99% de confianza que el modelo de recomendaciones

personalizado tiene una mayor tasa de conversión que las apuestas realizadas por los product manager.

Con los datos disponibles también es posible analizar las ventas generadas solo en las líneas recomendadas (vestuario hombre, vestuario mujer, vestuario juvenil hombre, vestuario juvenil mujer, calzado hombre y calzado mujer), con la diferencia que en este caso la conversión no se puede analizar en base a las órdenes, sino que a la cantidad de productos vendidos (una orden puede tener más de un producto). Los resultados obtenidos se pueden ver en la tabla a continuación.

	Aperturas	Ventas	Artículos vendidos	ARPU	Conversión
Control	29.934	\$988.880	44	\$33	0,15%
Tratamiento	29.521	\$1.392.170	76	\$47	0,26%
Lift	-	41%	73%	43%	75%

Tabla 36- Lift de ventas. Fuente: Elaboración propia

En estos resultados específicos a vestuario y calzado, es posible notar que el lift del monto vendido y el ARPU, son mayores que en la tabla 35, este comportamiento se cree que es debido a que, al analizar los montos vendidos en todas las líneas agrupadas se diluye el efecto del email al tener vestuario un ticket más bajo que otras líneas como electro y dormitorio. Con respecto al ticket promedio, nuevamente se da el caso de que el grupo de tratamiento tiene uno menor (22.474 v/s 18.318), corroborando así que las recomendaciones realizadas con el modelo personalizado provocan venta de productos más baratos. Por último, para la tasa de conversión, nuevamente se realiza un test z de proporciones con dos colas, que da como resultados un estadístico de 3,0004 y un p-valor de 0,0027, por lo tanto, es posible rechazar con un 95% de confianza la igualdad de proporciones entre el grupo de control y tratamiento. Luego, realizando un test de una cola para verificar que el grupo de tratamiento tenga una mayor tasa de conversión, se tiene un estadístico z de 3.0004 y un p-valor de 0.00135, por lo tanto, con un 99% de confianza es correcto señalar que, aislado a líneas de vestuario y calzado, el modelo de recomendaciones personalizado tiene una mayor tasa de conversión mayor que las apuestas realizadas por los product manager.

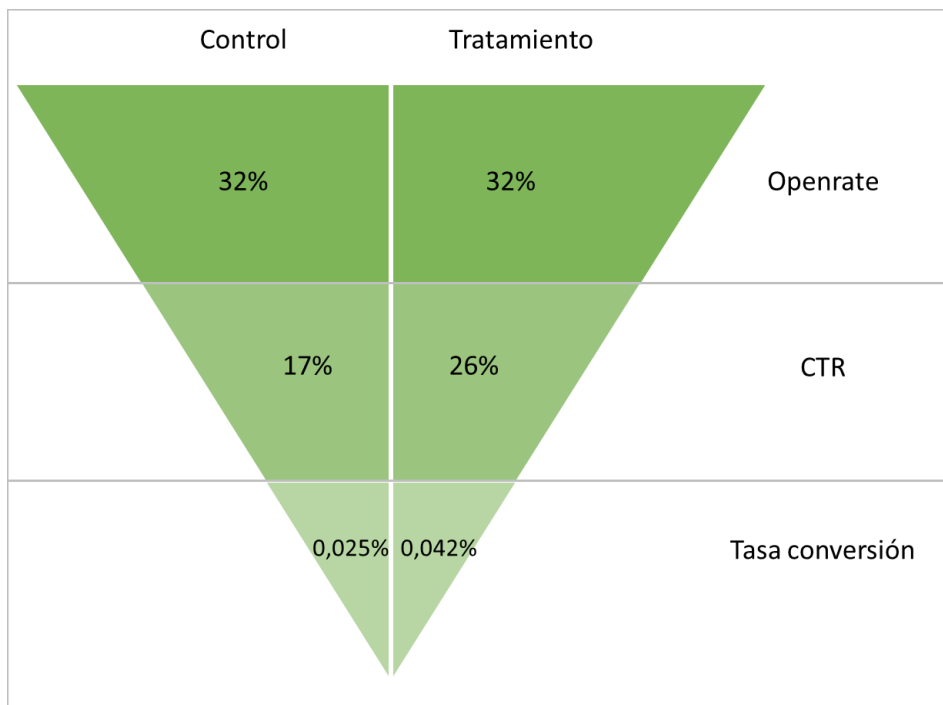


Ilustración 4-Funnel conversión experimento. Fuente: Elaboración propia

Recopilando los resultados experimentales es posible generar el funnel de la ilustración 4, donde se ve que las recomendaciones personalizadas generan mayor tráfico al sitio web que las no personalizadas, además, de secciones anteriores se sabe que este tráfico es sobre elementos de mayor interés en el correo. Por otro lado, en el ámbito de las ventas, se tiene que el grupo de tratamiento tiene mayor tasa de conversión, pero menor ticket promedio, aun así, el balance total de ventas a favor de las recomendaciones personalizadas con un mayor ARPU.

7. Conclusiones

La conclusión más importante corresponde a que los sistemas de recomendación basados en filtros colaborativos entregan mejores resultados empíricos que recomendaciones generales obtenidas en base al criterio de juicio de experto, éstas mejoras son reflejadas en un aumento de un 75% en la conversión y un 41% en ventas (no medida la significancia) dentro de las líneas recomendadas. Además, en específico el modelo de factorización de matrices implícito implementado tiene un bajo costo de ejecución, que alcanza los 6 minutos en un computador local para 100.000 clientes y tiene la capacidad de escalar a medida que se aumenta el número de clientes, por lo tanto, no solo las recomendaciones personalizadas resultan ser efectivas en cuanto a la conversión y clics, sino que además es barata en su ejecución. Por otro lado, como punto en contra, se tiene que solo puede ser implementado en clientes habituales del negocio, ya que el mínimo de 5 transacciones solicitado deja fuera más del 60% del conjunto de clientes utilizado para seleccionar la muestra. Es por ello, que en caso de ser aplicado, se recomienda utilizarlo junto otras estrategias para poder alcanzar el total de los clientes. Por ejemplo, la técnica que recomienda los productos más vendidos durante el mes anterior es una metodología que podría ser aplicada a todos los clientes.

Con respecto a los modelos en particular, se tiene que KNN entrega los mejores resultados en cuanto a las recomendaciones, pero su alto costo de ejecución y la poca escalabilidad del algoritmo en base a fuerza bruta, impide que se pueda implementar en un ambiente operacional. Se propone buscar otras alternativas de implementación para la búsqueda de vecinos, ya sea métodos más sofisticados o una implementación en entornos con mayor capacidad de procesamiento como lo es Google Cloud. Por otro lado, se sugiere generar fuentes de datos estructurados con las características detalladas de los productos ofrecidos en tienda y página web. De este modo, sería posible probar metodologías de filtros en base a contenido y ver si hay mejoras al usarlos de manera individual o en conjunto con recomendaciones de filtros colaborativos.

Como propuestas para el negocio, se tiene que las recomendaciones personalizadas pueden implementarse en múltiples puntos de contacto con el cliente como: campañas de email marketing, sugerencias en la página web, alimentando los cupones de descuento que se entregan en las cajas al realizar una compra y en las campañas de descuento de click and collect. Una estimación en base a los resultados obtenidos aplicados a las campañas de email marketing y el segmento acotado de clientes, entrega que el uso de la metodología permitiría alcanzar un aumento total de un 24% en la venta incremental de las campañas.

Por último, se concluye que una aplicación interesante podría ser evaluar los resultados que se den en otros negocios del holding que tengan la misma dificultad de recomendar productos dentro de un alto volumen de estos, ya que las metodologías son lo suficientemente versátiles para poder ser replicadas en otras tiendas. En particular, potenciales usuarios del modelo de recomendación pueden ser las tiendas de mejoramiento de hogar, supermercados y la nueva plataforma de Marketplace.

8. Bibliografía

1. Pathak, B., Garfinkel, R., Gopal, R. D., Venkatesan, R., & Yin, F. (2010). Empirical analysis of the impact of recommender systems on sales. *Journal of Management Information Systems*, 27(2), 159-188.
2. Lee, D., & Hosanagar, K. (2014). Impact of recommender systems on sales volume and diversity. *Thirty Fifth International Conference on Information Systems*, Auckland 2014, 880-895.
3. Mackenzie, I., Meyer, C., & Noble, S. (s.f). How retailers can keep up with consumers. [en línea] disponible en: <https://www.mckinsey.com/industries/retail/our-insights/how-retailers-can-keep-up-with-consumers> [visto el 21 Abr. 2018].
4. Inside Retail Asia. (2018). How Alibaba uses artificial intelligence to change the way we shop - Inside Retail Asia. [en línea] disponible en: <https://insideretail.asia/2017/06/07/how-alibaba-uses-artificial-intelligence-to-change-the-way-we-shop/> [visto el 21 Abr. 2018].
5. Isinkaye, F. O., Folajimi, Y. O., & Ojokoh, B. A. (2015). Recommendation systems: Principles, methods and evaluation. *Egyptian Informatics Journal*, 16(3), 261-273.
6. Jannach, D. (2012). Part I: Introduction to basic concepts. In *Recommender systems: An introduction*, 13-206. New York: Cambridge University Press. Libro
7. Herlocker, J. L., Konstan, J. A., Borchers, A., & Riedl, J. (1999, August). An algorithmic framework for performing collaborative filtering. In *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 230-237). ACM.
8. Leskovec, J., Rajaraman, A., & Ullman, J. D. (2014). *Mining of massive datasets*. Capítulo 9. Cambridge university press.
9. Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix factorization techniques for recommender systems. *Computer*, (8), 30-37.
10. Hu, Y., Koren, Y., & Volinsky, C. (2008, December). Collaborative filtering for implicit feedback datasets. In *Data Mining, 2008. ICDM'08. Eighth IEEE International Conference on* (pp. 263-272). Ieee.
11. Gershman, A., Meisels, A., Lüke, K. H., Rokach, L., Schclar, A., & Sturm, A. (2010, June). A Decision Tree Based Recommender System. In *IICS* (pp. 170-179).
12. Ben Frederickson. (s.f). Implicit documentation. [en línea] disponible en: <https://implicit.readthedocs.io/en/latest/index.html> [visto el 12 Oct. 2018].

9. Anexos

9.1. Anexo A: Matriz con compras de clientes

Clase-marca	Clientes enmascarados					
	5662	11190	12822	13598	14187	14232
CAMISA ML-VANDINE				1		
CAMISA SP ML-MP (3)	1					
CASACA CUERO – MP (2)				1		
CINTURON-HFRANK			1			
GORROS- MP(5)					1	
JEANS BASICOS- MP (6)						2
JEANS MODA-MP (2)					2	
JEANS MODA-MP (7)					1	
PANTALON DEP HOMBRE- ADIDAS	1					
PANTALON-MP (1)				1		
PANTALON-DOCKERS		4				
PANTALON-MP (3)	1					
PANTALON-PRABANNE			1			
PANTALON-MP (4)	1					
PARKAS-MP (5)		2				
POLERA MC-MP (8)			1		3	
POLERA MC- MP (2)				1		
POLERA MC-MAUI						1
POLERAS- MP (1)			1		2	
POLERAS- MP (3)	1					
POLERAS- MP(4)	1					1
POLERON HOMBRE-K2						
RELOJES ANTEOJOS Y VARIOS- MP(2)					1	

SWEATER ML SPORT-DOCKERS		2				
SWEATER ML SPORT-MP (4)						1
TOPS DEP HOMBRE-ADIDAS						2
ZAP FOOTBALL HOM-NIKE				1		
ZAP FOOTBALL HOM-PUMA					1	
ZAP OUTDOOR HOM-ADIDAS		1			1	
ZAPATOS HOMBRE-MP (8)	1					
ZAPATOS HOMBRE-MP (1)			2			
ZAPATOS HOMBRE-GUANTE		2				
CONTEO DE INTERACCIONES	7	5	5	5	8	5

9.2. Anexo B: Guess segmento muy alto

BOTTOMS SEÑORA-JEANS		
Marca	Precio promedio	Segmento
GUESS	119000	3
SANCTUARY	41900	2
PRIVILEGE	32790	2
CANADIENNE	17862	1
MADREMIA	23325	1
APOLOGY	18430	1
MA.DE	22787	1
DI PONTTI	4356	0
MARCA PROPIA	14251	0
MARCA PROPIA	14596	0
ECLIPSE	12465	0
CURVI	13088	0