

Implementación de modelo de aprendizaje automático para el ordenamiento de pasillos en un supermercado digital basado en el contenido del carrito de compras del usuario

Daniela Ramos García, Raymundo Guzmán Mata, Luis Antonio Barajas Ramirez, Sergio Ortiz Malpica, Arturo Durán Castillo

Filiación

Tecnológico de Monterrey

Introducción

El comercio electrónico de comestibles ha experimentado un crecimiento exponencial en los últimos años. Con el crecimiento del comercio electrónico y la evolución de los hábitos de consumo, se vuelve cada vez más importante entender cómo las personas compran, qué productos combinan frecuentemente y cómo se mueven dentro de una tienda. Esto ha planteado nuevos desafíos en términos de usabilidad y eficiencia de las interfaces de compra en línea. En plataformas con diseños estáticos de pasillos, los usuarios deben realizar desplazamientos (scrolls) excesivos para encontrar productos, lo que incrementa el “esfuerzo de navegación” y puede derivar en tasas de abandono de carrito elevadas (YasserH, 2017).

Las técnicas de recomendación basadas en patrones secuenciales han logrado avances notables en la predicción del siguiente ítem de compra: tomando como referencia los artículos más relevantes sobre el análisis de datos para optimización de comportamiento de páginas web para compras en línea, destacan modelos que se orientan principalmente a sugerir productos y no exploran la optimización dinámica del layout de la interfaz.

En este trabajo proponemos un clasificador supervisado que, a partir del contenido parcial del carrito y covariables contextuales (p. ej., hora del día, historial de compras), predice en tiempo real el siguiente pasillo que visitará el usuario y reordena dinámicamente la presentación de secciones en la tienda en línea. Definimos el esfuerzo de navegación como la suma de las distancias absolutas entre posiciones de pasillos según el orden en que aparecen los pasillos dentro del layout de la página y evaluamos su reducción mediante simulaciones comparativas entre layouts estáticos y dinámicos. Los resultados preliminares demuestran que un layout adaptativo puede disminuir significativamente el esfuerzo de navegación, mejorando así la eficiencia y satisfacción en la experiencia de compra en línea.

Marco Teórico

Actualmente, los supermercados digitales presentan una organización estática de sus pasillos virtuales, lo que obliga al usuario a recorrer listas largas y poco personalizadas para completar su compra. Este diseño no considera el contexto individual del usuario ni sus hábitos de compra, lo cual incrementa el esfuerzo de navegación y puede afectar la tasa de conversión. El problema puntual que se aborda en esta investigación es la ineficiencia en la organización de los pasillos

dentro de un supermercado en línea, que no se adapta dinámicamente a las decisiones del usuario en tiempo real. En nuestra solución definimos que el esfuerzo de navegación será medido como la suma de las distancias verticales (en número de scrolls o posiciones) entre pasillos consecutivos en el orden en que el usuario va agregando productos al carrito, dentro de un mismo departamento. Para ello, se asignará un número de orden a cada pasillo dentro del layout, y por cada producto agregado, se calculará cuántos "saltos" hubo que hacer para pasar del pasillo anterior al actual. Esta métrica permite cuantificar el desplazamiento necesario durante la navegación y sirve como base para comparar el layout estático (baseline) con el layout dinámico sugerido por el modelo.

Existen dos enfoques clave en esta investigación: la *minería de datos* aplicada al análisis de cestas de mercado y la *optimización dinámica del layout* de tiendas, siendo más relevante el segundo enfoque. La minería de datos es el proceso de buscar patrones, relaciones o conocimientos útiles en grandes conjuntos de datos. La minería de datos ha sido ampliamente utilizada en retailers (tiendas virtuales como Amazon o físicas como Walmart) para identificar patrones de compra frecuentes a través de técnicas como las reglas de asociación (las reglas de asociación identifican qué productos suelen comprarse juntos). El algoritmo Apriori, por ejemplo, permite detectar productos que suelen comprarse juntos, lo cual ha servido tradicionalmente para sugerencias de compra o diseño de promociones.

En el contexto de optimización dinámica del layout, diversos enfoques han sido propuestos para modelar las interacciones entre usuarios y productos durante una sesión. Cuando hablamos de optimización dinámica del layout en el contexto de este problema, necesitamos entender que son los modelos de aprendizaje automático. El *aprendizaje automático* es una rama de la inteligencia artificial que permite a las computadoras

aprender de los datos para tomar decisiones. En la implementación de nuestra solución el uso de aprendizaje automático (ML) es necesario porque permite modelar el comportamiento de compra a partir del historial de datos, detectar patrones de transición entre productos y anticipar las necesidades del usuario. Este tipo de predicción no es posible mediante reglas estáticas o enfoques tradicionales, ya que requiere adaptabilidad continua a las acciones del usuario durante la sesión. A pesar de que existen sistemas de recomendación en e-commerce, no se ha aplicado con suficiente profundidad el reordenamiento dinámico del layout de pasillos a nivel de interfaz según el contenido del carrito, lo que representa una oportunidad de innovación y diferenciación.

Entre este tipo de modelos (como mencionado de manera general en la introducción) destacan los modelos basados en factorización matricial, cadenas de Markov (Rendle, Freudenthaler, & Schmidt-Thieme, 2010), y redes neuronales recurrentes para modelar dinámicamente las interacciones de sesión (Hidasi, Karatzoglou, Baltrunas, & Tikk, 2015). Empezamos por los modelos basados en factorización matricial. Estos modelos permiten predecir productos de interés para cada usuario a partir de patrones latentes en sus historiales de compra. Este método se usa mucho en sistemas que hacen recomendaciones, porque puede adivinar qué productos podrían gustarle a una persona, aunque no haya comprado mucho antes, comparando sus gustos con los de otros usuarios parecidos. Seguido de esto, tenemos a las cadenas de Markov, que modelan la probabilidad de transición entre productos, considerando únicamente el estado actual o reciente del usuario. Esto resulta útil para anticipar la siguiente acción de compra, simplificando el problema sin requerir el análisis de toda la secuencia previa. Finalmente, las redes neuronales recurrentes (RNN) pueden entender relaciones más complicadas entre los productos que el usuario va eligiendo, ya que tienen una especie de memoria que recuerda todo lo

que ha ido agregando durante su compra. Nuestra solución no toma en cuenta la estética y diseño de la página sino que la experiencia de compra.

Planteamiento formal/técnico del problema

La experiencia de compra en línea no solo abarca la interfaz de usuario, si no también implica mucho en conocer el cliente y su comportamiento a la hora de realizar una compra. En la actualidad, los supermercados enfrentan el desafío de mejorar la experiencia de compra de sus clientes mientras optimizan sus operaciones internas. Con el crecimiento del comercio electrónico y la evolución constante de los hábitos de consumo, el reto de optimizar este proceso de compra en el usuario se ha vuelto más grande. Esto exige un mayor entendimiento de los patrones de compra, de las combinaciones frecuentes de productos y el comportamiento del cliente dentro de la tienda.

Esta investigación se propone aplicar técnicas avanzadas, combinándolas con métodos de optimización dinámica del layout de tienda, con el fin de generar una propuesta que permita rediseñar la organización de los pasillos, dado un patrón de compra, de este modo reduciendo los tiempos de recorrido y aumentando la satisfacción del cliente.

La relevancia de este trabajo radica en que contribuye tanto a la mejora de la eficiencia operativa como a la personalización de la experiencia de compra, generando beneficios económicos mediante decisiones informadas basadas en datos de utilidad. Asimismo, la motivación del estudio se fundamenta en la posibilidad de adaptar y validar modelos ya existentes en nuevos entornos, integrando enfoques recientes como algoritmos evolutivos, mecanismos de seguimiento de usuarios y modelos predictivos de comportamiento, lo cual ofrece una oportunidad única para innovar en la gestión de espacios comerciales físicos y digitales.

Objetivos de la investigación

General

Desarrollar y validar un modelo de aprendizaje automático que, a partir del contenido del carrito de

compras de un usuario y el histórico de pedidos realizados en la aplicación, prediga los pasillos con mayor probabilidad de ser visitados, esto con el fin de organizar los pasillos dentro de cada departamento con el propósito de minimizar el esfuerzo de navegación del comprador y mejorar la eficiencia de llenado de carritos.

Específicos

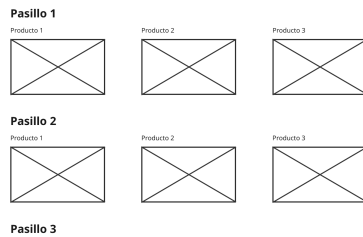
1. Explorar el conjunto de datos: Analizar la estructura y calidad de los archivos proporcionados, además de identificar valores faltantes e inconsistencias que puedan afectar al modelado.
2. Definir la métrica de esfuerzo de navegación: Formalizar un algoritmo que calcule el esfuerzo basado en la posición vertical de cada pasillo y el número de desplazamientos necesarios para acceder a él.
3. Diseñar y entrenar el modelo de ranking de pasillos: Seleccionar los algoritmos adecuados a implementar.
4. Integrar el modelo para su validación: Desarrollar una simulación en la que se reordenen los pasillos en tiempo real cada vez que el carrito se actualice para posteriormente medir el esfuerzo para realizar los pedidos con el nuevo ordenamiento.
5. Comparar el desempeño frente a otros acomodos: Cuantificar la reducción de esfuerzo promedio respecto a ordenamientos aleatorios con el layout propuesto. Aplicar pruebas para confirmar la significancia de las mejoras.

Metodología formal

La metodología propuesta tiene como objetivo validar que el reordenamiento de pasillos reduzca el esfuerzo que realiza un usuario al llenar su carrito de compras en línea. Como primer paso, es necesario definir la manera en que mediremos el esfuerzo requerido por un usuario para llenar su carrito de compras. Así que, analizando el sitio web de InstaCart, observamos que cada página de la aplicación representa a un departamento, dentro de un departamento se

encuentran, enlistados verticalmente, los pasillos que pertenecen a este. Así mismo cada pasillo cuenta con una lista horizontal de sus respectivos productos.

Departamento 1



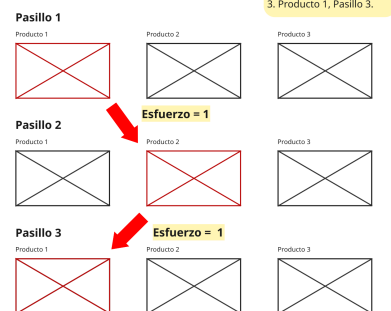
De esta forma, tomando en cuenta el histórico de pedidos realizados y a su vez, el orden en el que fueron agregados al carrito los productos comprados, realizamos el siguiente algoritmo:

1. El usuario, al entrar a un departamento, se encuentra en la posición inicial ($p = 0$).
2. El usuario necesita desplazarse n pasillos hasta llegar al que contiene el producto que se agregó al carrito. La posición vertical en la que se encuentra el pasillo es la cantidad de esfuerzo que necesita realizar para llegar a este.
3. Si el siguiente pasillo que se visita se encuentra en otro departamento, entonces el usuario tiene que entrar a esta página, por lo que su posición inicial vuelve a 0, por otro lado, si el siguiente pasillo se encuentra dentro del mismo departamento, entonces calculamos la diferencia absoluta entre la posición del pasillo en el que se encuentra actualmente y la posición del pasillo a visitar.

En la información proporcionada, contamos con los pasillos que pertenecen a cierto departamento, sin embargo, no contamos con el orden en el que aparecen en la página. Así que, con el propósito de comparar el layout que generará nuestro modelo de Machine Learning, asignamos posiciones aleatorias a los pasillos dentro de los departamentos. La intención es iterar sobre la generación de múltiples layouts

aleatorios y para cada uno de ellos aplicar el algoritmo de cálculo de esfuerzo con el fin de observar cómo se comportan los datos en diferentes escenarios. De esta forma, podremos tener una distribución de esfuerzo bajo condiciones no optimizadas, lo cual nos permitirá establecer una línea base para comparar el desempeño de nuestro layout propuesto.

Departamento 1



Productos agregados:
1. Producto 1, Pasillo 1.
2. Producto 2, Pasillo 2.
3. Producto 1, Pasillo 3.

Tomando en cuenta esto, nosotros podemos identificar las siguientes variables para nuestro análisis y planteamiento de nuestra solución:

Variable dependiente.

Esfuerzo del usuario: medido como la suma de las distancias absolutas entre pasillos dentro de un mismo departamento, en el orden en que se agregan productos al carrito.

Variable de tratamiento.

Reordenamiento dinámico de pasillos:

- Valor 0: el orden de los pasillos es estático (baseline).
- Valor 1: el sistema reorganiza los pasillos en tiempo real según los productos en el carrito.

Covariables.

- Hora del día: mañana, tarde o noche (puede influir en los hábitos de compra).
- Número total de productos en la compra: más productos podrían implicar más esfuerzo.
- Tipo de usuario: nuevo vs. recurrente.

- Categoría de productos seleccionados: algunos departamentos pueden tener layouts más complejos.
- Frecuencia de compra de los productos: productos comunes vs. raros.

Algoritmo de Machine Learning

En nuestro caso, el esfuerzo de los usuarios se mide utilizando un algoritmo específico que calcula la cantidad de movimiento necesario para recolectar los productos en el carrito, considerando el orden en que se agregan y la disposición de los pasillos. El algoritmo recorre cada pedido, ordenando los productos según el orden en que fueron añadidos al carrito (“add_to_cart_order”), y suma la distancia recorrida entre la posición actual del usuario y la ubicación del siguiente producto, reiniciando la posición cada vez que cambia de departamento.

Para comparar el impacto de nuestro algoritmo de reordenamiento dinámico de pasillos frente al ordenamiento aleatorio, aplicamos este cálculo de esfuerzo a ambos escenarios. La variable de tratamiento (“reordenamiento dinámico de pasillos”: 0 para estático, 1 para dinámico) se utiliza para diferenciar los grupos. Posteriormente, analizamos y comparamos los valores de esfuerzo obtenidos en cada grupo.

En este contexto, el modelo de machine learning más adecuado sería una regresión lineal, ya que el esfuerzo es una variable numérica continua. Este modelo nos permitirá cuantificar la reducción promedio del esfuerzo atribuible al reordenamiento dinámico, controlando por la variable de tratamiento. Así, podremos demostrar de manera objetiva y estadística que nuestro algoritmo reduce el esfuerzo requerido por los usuarios en comparación con un layout aleatorio o estático.

Pruebas y experimentos iniciales

Con el fin de validar la viabilidad de nuestra hipótesis y sentar las bases para el desarrollo del modelo de

machine learning, se realizaron pruebas preliminares utilizando un subconjunto del dataset de Instacart. Estas pruebas tienen como objetivo observar si el reordenamiento dinámico de pasillos, basado en los productos que el usuario va agregando al carrito, puede efectivamente reducir el esfuerzo de navegación durante la compra. A continuación, se detallan las variables consideradas y los resultados obtenidos en esta fase inicial.

Para dar inicio a la experimentación, se generará una columna adicional en el dataset resultante del merge entre todo los datasets. Esta nueva columna representará el número de orden asignado a cada pasillo dentro de su respectivo departamento. Inicialmente, este orden será aleatorio, simulando un layout base sin optimización. A partir de esta estructura, se analizarán múltiples sesiones de compra simuladas, observando cómo los usuarios se desplazan entre pasillos y departamentos conforme agregan productos al carrito. Este enfoque permitirá medir el esfuerzo de navegación en distintos escenarios de ordenamiento y servirá como base para comparar el impacto de estrategias de reordenamiento dinámico en la experiencia de compra.

Este nuevo dataset enriquecido, que incluye la columna con el orden asignado a cada pasillo dentro de su departamento, también servirá como base para entrenar un modelo de aprendizaje supervisado de clasificación. En este contexto, cada fila del dataset puede representar una instancia en la que el sistema debe decidir la posición óptima de un pasillo, dadas ciertas condiciones como la hora del día, los productos ya agregados al carrito, el departamento al que pertenece el pasillo y la frecuencia histórica de compra. La variable objetivo será la posición ideal del pasillo (aisle_order), mientras que las variables predictoras incluirán tanto características contextuales como comportamentales. De esta manera, el modelo podrá aprender patrones que permitan anticipar qué pasillos

deben mostrarse primero para reducir el esfuerzo del usuario, y así optimizar dinámicamente la experiencia de navegación en el supermercado en línea.

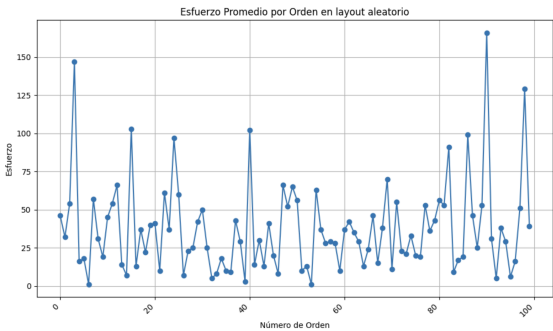
Resultados.

Generando 20 layouts aleatorios para los pasillos y utilizando las órdenes de productos ya existentes que fueron provistas, realizamos los cálculos de los esfuerzos.

Promedio de orden con layout aleatorio: 37.52

order_id	esfuerzo
0	2
1	3
2	4
3	5
4	6
...	...
95	103
96	104
97	105
98	106
99	107

El esfuerzo mostrado por cada order_id, es el promedio de esfuerzo que se tomó en los 20 diferentes layouts.



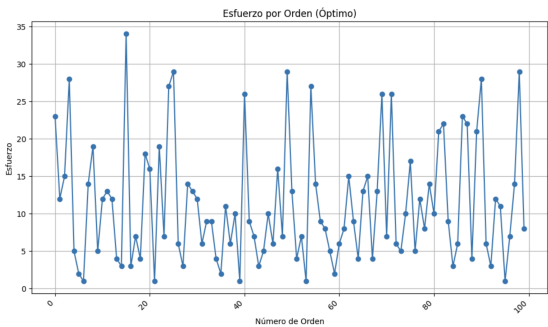
Analizando individualmente el promedio de esfuerzo por cada orden, podemos ver que se tienen grandes cantidades de esfuerzos siendo el máximo, aproximadamente de 170.

Continuando, realizamos un cambio en la forma de calcular el esfuerzo para realizar un layout óptimo, que será similar a nuestro sistema dinámico. Utilizamos las mismas órdenes de producto y un layout aleatorio. Estos fueron los resultados obtenidos:

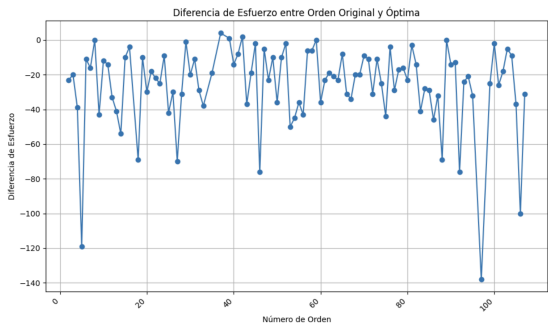
Promedio de orden con layout optimo: 11.3

order_id	esfuerzo
0	2
1	3
2	4
3	5
4	6
...	...
95	103
96	104
97	105
98	106
99	107

El promedio de esfuerzo total se redujo 26.22 puntos en comparación a los layouts aleatorios.



El esfuerzo máximo fue 34 puntos, reduciendo un casi 80% del máximo de layouts aleatorios (166 puntos).



Para poder comparar los resultados de una mejor manera, realizamos la diferencia entre esfuerzo de layout óptimo vs esfuerzo layout aleatorios. Estos resultados obtenidos con los experimentos iniciales muestran una mejora significativa en el desempeño de esfuerzo. Específicamente, se observó que el promedio de esfuerzo total se redujo en -26.22 puntos,, lo que indica una optimización considerable en la eficiencia del proceso. La máxima diferencia de esfuerzos fue de -138 puntos y 4 la minina.

El cálculo y layout óptimo se realiza casi de la misma manera que los aleatorio, solo que se modifica en una sección:

El bloque *if(prev_dep != current_dep)*: sirve para identificar el momento en que el usuario pasa de un departamento a otro dentro del supermercado.

Cuando esto ocurre, se actualizan dos variables clave: *prev_dep* (el departamento anterior) y *actual_user_position* (la posición actual del usuario en el pasillo). Además, se suma 1 al esfuerzo, que representa el esfuerzo mínimo necesario para entrar a un nuevo departamento y seleccionar el primer producto allí. Esto es fundamental ya que en la vida real, cambiar de departamento implica un esfuerzo.

Por eso, cada vez que el usuario entra a un nuevo departamento, el esfuerzo se aumenta en 1, y la posición del usuario se actualiza al pasillo donde se encuentra el primer producto de ese departamento.

El modelo simplifica el recorrido: cada vez que se cambia de departamento, el usuario “salta” directamente al pasillo necesario del nuevo departamento, y desde ahí continúa su recorrido. Esta es una forma de calcular el esfuerzo mínimo posible, ignorando la distancia real entre departamentos y enfocándose solo en los movimientos dentro de cada departamento.

Es importante resaltar que este layout y cálculo óptimo es una prueba en donde se buscaba enfatizar que al optimizar el diseño de layout dinámico, se tiene un impacto positivo y reducción en el esfuerzo del usuario.

Referencias.

- Arboleda, F.J.M., Garani, G., Correa, A.F.A. Supermarket Product Placement Strategies Based on Association Rules (2024) IAENG International Journal of Computer Science, 51 (6), pp. 650-662.
- Saberi, Z., Hussain, O. K., & Saberi, M. (2023). Data-driven personalized assortment optimization by considering customers' value and their risk of churning: Case of online grocery shopping. *Computers & Industrial Engineering*, 184, 109328. <https://doi.org/10.1016/j.cie.2023.109328>
- Liashenko, O., & Yakymchuk, B. (2022). Modeling the Optimal Grocery Store Trading Area Using Machine Learning Methods. *Proceedings of the Information Technology and Implementation (IT&I-2022)*, CEUR Workshop Proceedings, 3347, 325–331. https://ceur-ws.org/Vol-3347/Short_3.pdf
- Boyer, K. K., & Hult, G. T. M. (2005). Customer behavior in an online ordering application: A decision scoring model. *Decision Sciences*, 36(4), 569–602. <https://doi.org/10.1111/j.1540-5414.2005.00103.x>
- Pillai, J., Vyas, O.P., Muyebe, M. (2013). HURI – A Novel Algorithm for Mining High Utility Rare Itemsets. In: Meghanathan, N., Nagamalai, D., Chaki, N. (eds) *Advances in Computing and Information Technology*. Advances in Intelligent Systems and Computing, vol 177. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-31552-7_54
- Tarigan, U., Tarigan, U. P. P., Rahman, I. H., & Rizkya, I. (2018). Design of facility layout with lean service and market basket analysis method to simplification of service process in the supermarket. *MATEC Web of Conferences*, 197, 14006. <https://doi.org/10.1051/mateconf/201819714006>
- Joe, T., Sreejith, R., & Sekar, K. (2019). Optimization of store layout using market basket analysis. *International Journal of Recent Technology and Engineering (IJRTE)*, 8(2), 6459–6463. <https://doi.org/10.35940/ijrte.B2207.078219>
- García-Magariño, I., Lloret, J., & Pawar, P. (2021). An E-Business Event Stream Mechanism for Improving User Tracing Processes. *Computers, Materials & Continua*, 69(1), 1143–1158. <https://doi.org/10.32604/cmc.2021.014278>
- Budaraju, Raja Rao & Jammalamadaka, Sastry. (2025). Crowd Distance Induced Multi Objective Binary Salp Swarm Optimization Algorithm for Mining High Frequency and Utility Itemsets. *SN*

Computer Science. 6. 1-14.
10.1007/s42979-025-03725-8.

10. Gan, W., Lin, J. C., Fournier-Viger, P., Chao, H., Hong, T., & Fujita, H. (2018). A survey of incremental high-utility itemset mining. *WIREs Data Mining and Knowledge Discovery*, 8(2).
<https://doi.org/10.1002/widm.1242>