

Diseño e implementación de motor de afinidad para personalización comercial B2B en consumo masivo

Lic. Abril Noguera

Carrera de Especialización en Inteligencia Artificial

Director: Ing. Juan Pablo Rodríguez Varela (ITBA)

Jurados:

Jurado 1 (pertenencia)
Jurado 2 (pertenencia)
Jurado 3 (pertenencia)

Ciudad de Buenos Aires, diciembre de 2025

Resumen

En la presente memoria se describe el diseño e implementación de un motor de afinidad orientado a la personalización comercial en un entorno de negocio del sector de consumo masivo. Se desarrolló un sistema de recomendación capaz de estimar la relevancia de cada producto para cada cliente a partir de datos transaccionales, señales digitales y características contextuales, con el objetivo de generar listas priorizadas de sugerencias.

Para su desarrollo fueron fundamentales los conocimientos adquiridos en la carrera, tales como aprendizaje automático, aprendizaje profundo, validación de modelos, ingeniería de atributos y prácticas de MLOps para la trazabilidad y despliegue del sistema.

Agradecimientos

Esta sección es para agradecimientos personales y es totalmente **OPCIONAL**.

Índice general

Resumen	I
1. Introducción general	1
1.1. Marco de la propuesta	1
1.2. Definición del problema	2
1.3. Estado del arte	4
1.3.1. Referencias en sistemas de recomendación	4
1.3.2. Sistemas de recomendación en B2B	4
1.3.3. Caso de implementación	5
1.3.4. Lecciones aprendidas	5
1.4. Motivación	6
1.5. Objetivos y alcance	7
2. Introducción específica	9
2.1. Sistemas de recomendación	9
2.1.1. Funcionamiento de los sistemas de recomendación	9
2.1.2. Tipos de <i>feedback</i>	10
2.1.3. Filtrado colaborativo	10
2.1.4. Sistemas basados en contenido	11
2.2. Fuentes de información	11
2.3. Herramientas utilizadas	12
2.3.1. Plataformas de procesamiento distribuido	12
2.3.2. Gestión del ciclo de vida de modelos	12
2.3.3. bibliotecas de aprendizaje automático y profundo	12
2.3.4. bibliotecas de visualización	13
2.3.5. Control de versiones y colaboración	13
2.3.6. Consideraciones finales	13
3. Diseño e implementación	15
3.1. Diseño de solución	15
3.2. Preparación de los datos	16
3.3. Análisis exploratorio de los datos	17
3.3.1. Curvas de concentración de clientes y productos	17
3.3.2. Patrones de diversidad en el portafolio	19
3.3.3. Correlaciones entre variables transaccionales y digitales	21
3.3.4. Observaciones preliminares del análisis exploratorio	23
3.4. Ingeniería de atributos	23
3.4.1. Ventanas temporales de datos	24
3.4.2. Agregación de características	24
3.4.3. Codificación de variables categóricas	25
3.4.4. Tratamiento de valores atípicos	25
3.4.5. Normalización y escalado	26

3.4.6. Representación final de cliente-producto	27
3.5. Desarrollo de modelos	27
3.5.1. Filtrado colaborativo con ALS	28
3.5.2. Modelo híbrido con LightFM	29
3.5.3. Modelo de embeddings neuronales	29
3.5.4. Neural Collaborative Filtering (NCF)	30
3.6. Implementación	32
3.6.1. Diseño del pipeline de procesamiento	32
3.6.2. Integración con la infraestructura tecnológica	32
3.6.3. Estrategias de versionado y monitoreo	32
4. Ensayos y resultados	35
5. Conclusiones	37
Bibliografía	39

Índice de figuras

2.1. Ejemplo de representación de marcas de cerveza en un espacio de atributos.	10
3.1. Arquitectura de alto nivel del sistema de recomendación.	16
3.2. Concentración de productos en el portafolio.	18
3.3. Concentración de clientes.	18
3.4. Histograma de diversidad de portafolio: número de productos distintos por cliente.	19
3.5. Diversidad de portafolio segmentada por canal comercial.	20
3.6. <i>Log-log plot</i> de la popularidad de productos.	20
3.7. Mapa de calor de co-ocurrencia entre los 10 productos más relevantes.	21
3.8. Matriz de correlación entre variables transaccionales y digitales. . .	22
3.9. Correlación de variables con la compra en el mes siguiente.	22
3.10. Arquitectura del modelo <i>Neural Collaborative Filtering</i> híbrido. . . .	31

Índice de tablas

1.1. Ventajas y desventajas de enfoques en recomendación	6
1.2. caption corto	6
3.1. Tasa de recompra por combinación de señales	23

Dedicado a... [OPCIONAL]

Capítulo 1

Introducción general

Este capítulo tiene como propósito contextualizar el trabajo dentro del ámbito del consumo masivo y, en particular, de los modelos de negocio entre empresas (B2B). Se expone la relevancia que adquiere la personalización comercial en este sector y los desafíos que surgen al gestionar un portafolio amplio de productos frente a una base heterogénea de clientes. A partir de esta perspectiva se describe el problema central que motiva el desarrollo de un motor de afinidad y se señalan las limitaciones de los enfoques tradicionales de recomendación en entornos de alta variabilidad y escasez de datos.

Asimismo, se realiza una revisión introductoria de los principales sistemas de recomendación y de sus alcances en diferentes contextos, donde se destacan las particularidades que distinguen al escenario de negocio entre empresas. Finalmente, se presentan la motivación, la relevancia y los objetivos del trabajo, con el fin de ofrecer al lector una visión clara del problema abordado, de la importancia de su resolución y del recorrido que seguirá la memoria en los capítulos posteriores.

1.1. Marco de la propuesta

La industria del consumo masivo constituye uno de los motores más importantes de la economía, caracterizada por un volumen elevado de transacciones, la alta frecuencia de compra y la amplia variedad de productos que la conforman. La magnitud de este sector, junto con la fuerte competencia existente, obliga a las compañías a buscar permanentemente mecanismos que les permitan diferenciarse y mejorar la relación con sus clientes.

En este entorno, la relación comercial se establece entre una empresa proveedora y una red extensa de clientes minoristas que funcionan como canales de llegada al consumidor final. Estos clientes presentan una gran diversidad en cuanto a tamaño, ubicación geográfica, recursos disponibles y patrones de demanda. La heterogeneidad de la red de distribución genera que cada establecimiento tenga necesidades distintas y reaccione de manera diferente frente a la oferta de productos. Bajo estas condiciones, una estrategia comercial homogénea resulta insuficiente, ya que no logra capturar las particularidades de cada cliente ni ofrecerle productos que se ajusten de manera adecuada a su realidad.

La necesidad de personalización surge entonces como un factor estratégico central. Adaptar la oferta a las características específicas de cada cliente no solo incrementa la probabilidad de aceptación de los productos sugeridos, sino que también permite optimizar el uso del canal comercial, fortalecer la relación de largo

plazo y generar un impacto positivo en la eficiencia general del negocio. Las sugerencias ajustadas al contexto trascienden la idea de recomendar lo más vendido en términos absolutos: implica comprender la dinámica particular de cada cliente y priorizar aquellos productos que, dentro de un portafolio amplio, resulten más relevantes para su operación cotidiana.

A esta diversidad se suman factores que aumentan la complejidad del sector. La estacionalidad en la demanda, la influencia de promociones y campañas comerciales, la variabilidad en las preferencias de los consumidores finales y la constante rotación de productos dentro del catálogo configuran un escenario cambiante y difícil de predecir. La incorporación de artículos nuevos en el portafolio plantea, además, el desafío de la falta de contexto e información histórica que da perspectiva para guiar las recomendaciones.

En este marco, contar con herramientas que permitan personalizar la relación con cada cliente resulta indispensable. Un sistema capaz de priorizar los productos más relevantes para cada establecimiento aporta ventajas significativas: mejora la precisión de las recomendaciones, amplía la visibilidad de productos estratégicos, optimiza la gestión de los recursos comerciales y contribuye a consolidar vínculos más sólidos con los clientes minoristas. De esta manera, las recomendaciones a medida se convierten en un pilar fundamental para la sostenibilidad y la competitividad en el sector del consumo masivo.

1.2. Definición del problema

La empresa en la que se desarrolla este trabajo pertenece al sector del consumo masivo y opera bajo un modelo de venta directa a una red amplia y heterogénea de clientes minoristas. Esta red está compuesta por autoservicios, kioscos y comercios tradicionales distribuidos en todo el territorio nacional, lo que permite alcanzar una cobertura superior a los trescientos mil puntos de venta. La magnitud de esta operación, sumada a la diversidad de formatos y capacidades de los clientes, convierte a la personalización en una necesidad estratégica. A ello se suma la complejidad de un portafolio que incluye un gran número de marcas y presentaciones, lo que multiplica las posibles combinaciones cliente-producto y genera un desafío de gestión a gran escala.

El reto principal radica en estimar con precisión el interés que cada cliente podría tener en cada producto dentro del portafolio. Hoy en día, las decisiones comerciales se apoyan principalmente en el historial de ventas o en la popularidad general de los artículos, lo que conduce a una oferta relativamente homogénea. Este enfoque ignora las particularidades de los clientes y no captura la relevancia contextual de los productos. El problema central se expresa, entonces, en la ausencia de mecanismos que permitan calcular un nivel de afinidad entre cliente y producto capaz de reflejar con realismo el grado de interés que un artículo puede despertar en un punto de venta específico en un momento determinado.

Este desafío se ve amplificado por una serie de batallas que la empresa enfrenta de manera cotidiana en su estrategia comercial. La primera de ellas es la necesidad de pasar de un enfoque reactivo, basado en compras históricas, hacia una estrategia proactiva que permita anticipar tendencias de consumo y orientar la oferta en consecuencia. Para ello es indispensable contar con una herramienta

que adapte las recomendaciones de manera dinámica y alineada con el comportamiento observado en cada cliente.

Otra dimensión crítica es la optimización de recursos. La magnitud de la red comercial hace imposible abordar a todos los clientes con la misma intensidad, por lo que resulta fundamental identificar en qué productos y clientes concentrar los esfuerzos. Un motor de afinidad que jerarquice oportunidades de mayor impacto ofrece al equipo comercial la posibilidad de planificar visitas y diseñar ofertas más focalizadas, lo que mejora la eficiencia del canal.

La constante rotación del portafolio también representa un desafío de gran magnitud. Una proporción significativa de los productos se renueva cada año, lo que obliga a dar visibilidad a artículos sin historial de ventas y, al mismo tiempo, sostener el desempeño de categorías tradicionales. Este problema de arranque en frío limita la capacidad de los enfoques tradicionales para recomendar productos nuevos o poco frecuentes, lo que retrasa su incorporación en los puntos de venta y afecta el posicionamiento de la innovación en el mercado.

De manera similar, la inserción de nuevos clientes en la red sin historial de compras constituye un reto adicional. Cada semana se incorporan comercios que aún no cuentan con registros transaccionales suficientes para perfilar sus preferencias. Estos clientes suelen recibir sugerencias genéricas o basadas en promedios de segmentos, lo que reduce el atractivo de la oferta inicial y dificulta su integración temprana al canal digital. Una solución efectiva debería ser capaz de recomendar productos relevantes aun en ausencia de historial, al aprovechar señales contextuales y patrones de clientes similares.

La estacionalidad y las promociones constituyen otro factor de complejidad. La demanda de determinados productos fluctúa de manera pronunciada según la época del año o las campañas comerciales en curso. Un producto que en un período presenta alta relevancia puede perder vigencia en el siguiente, lo que provoca que reglas estáticas de recomendación queden rápidamente obsoletas. Para sostener la efectividad en este entorno dinámico se requiere un sistema flexible y capaz de adaptarse a variaciones temporales.

En conjunto, estos factores configuran un escenario donde la falta de personalización impacta de manera directa en los resultados del negocio. Sin un mecanismo que integre de manera sistemática los datos disponibles, se generan listas de productos poco relevantes para los clientes, se desperdician oportunidades de venta cruzada y se dificulta la adopción de innovaciones. Asimismo, el equipo comercial se ve limitado por información fragmentada, lo que reduce su capacidad de diseñar acciones específicas y de extraer valor de la gran cantidad de datos generados en el canal digital.

La solución propuesta apunta a superar estas limitaciones mediante el desarrollo de un motor de afinidad que calcule de forma periódica la relevancia de cada producto para cada cliente, y que integra señales transaccionales, interacciones digitales y atributos contextuales. Este motor tiene como objetivo generar rankings personalizados que orienten las recomendaciones tanto en el canal digital como en la gestión directa del equipo comercial. De esta forma, se busca avanzar hacia una estrategia más precisa, escalable y alineada con los objetivos de negocio, lo que habilita una gestión proactiva de portafolio y mejora la relación con los clientes de la red.

1.3. Estado del arte

El estado del arte permite ubicar este trabajo dentro de la evolución de los sistemas de recomendación. En esta sección se revisan los principales *benchmarks* en entornos B2C (del inglés, *Business to Customer*), los aportes de la literatura en contextos B2B y un caso de implementación en Brasil, para finalmente sintetizar los aprendizajes y señalar la brecha que orienta esta propuesta.

1.3.1. Referencias en sistemas de recomendación

El campo de los sistemas de recomendación se consolidó en los últimos veinte años como una de las áreas más dinámicas dentro de la inteligencia artificial aplicada. Sus desarrollos se originaron en entornos de consumo directo al público, donde el volumen de usuarios y la abundancia de señales digitales permitieron mejorar rápidamente la precisión y escalabilidad. A lo largo de este proceso, distintos hitos se transformaron en referencias obligadas y definieron *benchmarks* de la disciplina.

Uno de los puntos de inflexión fue el concurso Netflix Prize [1], que impulsó avances en factorización matricial y consolidó métricas de ranking como *recall* y *precision* en el análisis de desempeño. En paralelo, Amazon desarrolló un motor de recomendaciones basado en filtrado colaborativo *item-to-item*, reconocido por su capacidad de escalar en catálogos extensos y mantener robustez frente a grandes volúmenes de transacciones. MovieLens [2] se transformó en el dataset académico más utilizado, al servir como estándar para comparar algoritmos y validar resultados de manera consistente. Finalmente, plataformas como Spotify y YouTube llevaron la disciplina hacia modelos secuenciales y de aprendizaje profundo, capaces de personalizar en tiempo real a partir de interacciones en sesiones cortas.

Estos casos muestran cómo los sistemas de recomendación se convirtieron en el núcleo de la personalización digital y establecieron estándares en cuanto a precisión, escalabilidad y diversidad. Al mismo tiempo, reflejan un sesgo hacia contextos de B2C, donde las interacciones con consumidores finales son abundantes, explícitas y fácilmente trazables.

1.3.2. Sistemas de recomendación en B2B

En entornos de negocio entre empresas, la adopción de sistemas de recomendación es mucho más incipiente. La literatura identifica que, a diferencia de lo que ocurre en B2C, los procesos de compra en B2B suelen involucrar múltiples actores, ciclos de decisión más largos y una relación de largo plazo entre proveedor y cliente. Estas particularidades hacen que las soluciones desarrolladas para consumo final no se trasladen de forma directa.

El estudio presentado en [3] resalta el potencial de estas herramientas en B2B, al destacar que pueden reducir los costos de búsqueda, fortalecer vínculos comerciales y facilitar la introducción de productos en portafolios complejos. Sin embargo, también identifica desafíos clave: la necesidad de integrar datos dispersos de distintas fuentes, la importancia de la interpretabilidad para ganar confianza en decisiones de compra de alto valor y la dificultad de escalar modelos en contextos de menor densidad transaccional.

Si bien existe un reconocimiento académico del valor que los sistemas de recomendación pueden aportar en B2B, las implementaciones concretas son todavía escasas y carecen de estandarización. Esto genera una brecha significativa entre el potencial identificado y la práctica real, que representa una oportunidad de innovación para sectores como el consumo masivo.

1.3.3. Caso de implementación

Un antecedente particularmente relevante proviene de la propia organización, a través de la implementación de un sistema de recomendación en Brasil dentro de la plataforma digital BEES [4]. Este desarrollo tuvo como objetivo priorizar productos para cada punto de venta a gran escala, con el fin de reemplazar procesos manuales que en el pasado se realizaban en planillas y que resultaban poco eficientes.

El algoritmo principal implementado fue un filtrado colaborativo para feedback implícito, concretado mediante factorización matricial con el método *Alternating Least Squares* (ALS). El modelo utilizó como insumos tanto el historial de compras como señales digitales generadas en la aplicación, e incluyó búsquedas, visualizaciones de productos e interacciones con el carrito de compras. De este modo, se logró reducir sustancialmente la cantidad de recomendaciones enfocándolas en productos con mayor interés para el cliente, lo que marcó un avance significativo en la capacidad de personalizar la oferta a cada punto de venta.

Los resultados demostraron la viabilidad de este tipo de soluciones en un entorno B2B real y de gran escala. Sin embargo, también dejaron en evidencia limitaciones relevantes. La dependencia casi exclusiva del historial transaccional reforzó el problema del arranque en frío, tanto para productos recién incorporados como para clientes nuevos sin registros suficientes. Además, el sistema presentó limitaciones en diversidad de recomendaciones, ya que tendía a reforzar productos populares, y careció de un componente explícito para alinear los resultados con prioridades estratégicas de negocio.

El mismo documento identifica líneas de mejora hacia el futuro, como la incorporación de modelos híbridos que integren atributos de clientes y productos, el desarrollo de algoritmos de *clustering* para agrupar unidades de negocio con características similares y la inclusión de mecanismos que permitan diversificar resultados. Estas observaciones resultan especialmente valiosas para orientar el diseño de una solución adaptada al contexto argentino.

1.3.4. Lecciones aprendidas

El recorrido presentado permite extraer tres conclusiones principales. En primer lugar, los benchmarks internacionales muestran que los sistemas de recomendación son capaces de transformar industrias enteras cuando logran combinar precisión, escalabilidad y diversidad. En segundo lugar, la literatura sobre B2B reconoce la oportunidad de trasladar estos beneficios, pero también evidencia la falta de soluciones maduras que contemplen las particularidades de este tipo de relaciones comerciales. Finalmente, el caso de Brasil demuestra que es posible implementar un motor de recomendaciones en un contexto de consumo masivo B2B, pero también que persisten limitaciones en arranque en frío, diversidad y alineación con objetivos de negocio.

A modo de síntesis, la tabla 1.1 resume las ventajas y desventajas de cada uno de los enfoques revisados, e incluye la brecha identificada en el contexto argentino que motiva el desarrollo de un motor de afinidad adaptado a la realidad local. Este resumen permite enfatizar la necesidad de avanzar hacia un sistema que integre señales transaccionales y digitales, incorpore criterios estratégicos de negocio y se apoye en técnicas modernas de aprendizaje automático y profundo. El objetivo es superar las restricciones de los enfoques tradicionales y aportar un valor diferencial en la gestión comercial de la empresa en Argentina.

TABLA 1.1. Ventajas y desventajas de los enfoques revisados.

Enfoque / Caso	Ventajas principales	Desventajas principales
Benchmarks B2C (Netflix, Amazon, etc.)	Alta precisión y escalabilidad. Abundancia de datos y señales digitales. Estándares de evaluación consolidados.	Contextos con abundancia de <i>feedback</i> explícito/implícito, poco comparables al B2B. No consideran objetivos de negocio específicos.
Literatura B2B	Reconoce particularidades de clientes empresariales. Identifica beneficios en reducción de costos y fortalecimiento de relaciones.	Pocas implementaciones reales. Escasa estandarización de métricas y datasets. Desafíos de interpretabilidad y escalabilidad.
Caso Brasil (BEES)	Demostró viabilidad en gran escala. Integró compras e interacciones digitales. Mejora clara frente a procesos manuales.	Dependencia fuerte del historial transaccional (arranque en frío). Limitaciones en diversidad y alineación con objetivos estratégicos.
Brecha en Argentina	Oportunidad de adaptar aprendizajes globales y regionales. Potencial de integrar señales contextuales y digitales. Aplicación de técnicas modernas de aprendizaje automático y profundo.	Falta de solución probada en el contexto local. Mayor heterogeneidad y escala que en otros países.

TABLA 1.2. caption largo más descriptivo.

Especie	Tamaño	Valor
Amphiprion Ocellaris	10 cm	\$ 6.000
Hepatus Blue Tang	15 cm	\$ 7.000
Zebrasoma Xanthurus	12 cm	\$ 6.800

1.4. Motivación

La definición del problema mostró que la empresa enfrenta limitaciones para identificar con precisión qué productos resultan más relevantes para cada cliente

en cada momento, debido a factores como la rotación del portafolio, la estacionalidad de la demanda y la incorporación de nuevos clientes sin historial. El estado del arte, por su parte, evidencia que si bien existen avances notables en sistemas de recomendación y casos aplicados en entornos B2C, aún persiste una brecha en cuanto a soluciones robustas y adaptadas a escenarios B2B de consumo masivo.

La motivación de este trabajo surge de esa intersección: un problema claramente identificado en la operación local y un campo de conocimiento que ofrece enfoques valiosos pero todavía insuficientes para resolverlo en toda su complejidad. El diferencial de esta propuesta reside en integrar múltiples fuentes de información, transaccionales, digitales y contextuales, dentro de un motor de afinidad diseñado específicamente para el mercado argentino. Además, el trabajo incorpora la orientación explícita a objetivos de negocio y el uso de prácticas modernas de aprendizaje automático, aprendizaje profundo y MLOps, con el fin de garantizar escalabilidad, trazabilidad y alineación estratégica.

En este sentido, el trabajo no busca reproducir soluciones existentes, sino avanzar hacia un sistema que combine la rigurosidad técnica con la aplicabilidad práctica en un contexto desafiante, y que aporte un valor diferencial tanto en la gestión comercial de la empresa como en la evolución del conocimiento sobre sistemas de recomendación en consumo masivo B2B.

1.5. Objetivos y alcance

El propósito general de este trabajo es desarrollar un motor de afinidad que permita generar recomendaciones personalizadas de productos para cada cliente de la red de la empresa. El sistema se plantea como una herramienta capaz de integrar información transaccional, señales digitales y atributos contextuales con el fin de optimizar la gestión comercial, mejorar la efectividad de las sugerencias y facilitar la adopción de categorías estratégicas.

A partir de este objetivo general se desprenden metas específicas que orientan el desarrollo. En primer lugar, se busca analizar en detalle las fuentes de datos disponibles y transformarlas en insumos útiles para el modelado. Sobre esta base, se plantea la construcción de variables que reflejen el comportamiento de compra, las características de los productos y el contexto de cada cliente. Un segundo objetivo es implementar y comparar distintos enfoques de modelado, desde métodos de referencia hasta técnicas de factorización, modelos híbridos y arquitecturas profundas, con el objetivo de evaluar su desempeño con métricas de ranking como *recall@K*, *MAP@K*, cobertura y diversidad. De manera complementaria, se incluye la necesidad de diseñar estrategias que permitan afrontar el arranque en frío, tanto de productos recién incorporados al portafolio como de clientes nuevos sin historial de compras. Finalmente, se busca establecer un pipeline de entrenamiento y despliegue con prácticas de MLOps que garantice trazabilidad, reproducibilidad y escalabilidad del sistema.

El alcance del trabajo se limita a la construcción y evaluación de un prototipo funcional en un entorno controlado con datos reales de la empresa. Esto implica el análisis y preparación de la información, el desarrollo de modelos de recomendación y la evaluación de su desempeño a través de métricas definidas, e incluye escenarios de robustez frente a la incorporación de productos y clientes nuevos.

También, se contempla el diseño conceptual de la integración del motor con el canal digital y el apoyo al trabajo del equipo comercial.

Capítulo 2

Introducción específica

Este capítulo presenta los conceptos y componentes centrales que sustentan el trabajo. Se introducen los sistemas de recomendación y sus enfoques principales, se describen las fuentes de información empleadas y se detallan las plataformas y herramientas utilizadas para el procesamiento de datos, el modelado y la gestión de experimentos, que conforman la base tecnológica de la solución propuesta.

2.1. Sistemas de recomendación

Los sistemas de recomendación constituyen una de las aplicaciones más extendidas de la inteligencia artificial [5, 6], con un papel central en la reducción de la sobrecarga de información y en la optimización de decisiones de consumo. Su finalidad es generar sugerencias personalizadas que se ajusten a las características de cada cliente, lo que incrementa la relevancia de los productos ofrecidos y mejora la experiencia general de interacción con la empresa.

2.1.1. Funcionamiento de los sistemas de recomendación

El eje central del enfoque consiste en identificar relaciones de similitud entre productos, clientes o interacciones [5]. Estas relaciones pueden establecerse desde diferentes perspectivas. En primer lugar, es posible medir la similitud entre productos, lo que permite agrupar aquellos que suelen adquirirse en conjunto o que comparten atributos comunes. En segundo lugar, puede analizarse la similitud entre clientes, de modo que las preferencias observadas en un grupo con comportamientos semejantes permitan anticipar las elecciones de otros con perfiles cercanos. Por último, también resulta clave la similitud entre interacciones, que considera el historial de comportamientos de un cliente, como sus compras o búsquedas, para anticipar futuras decisiones.

Un ejemplo ilustrativo, representado en la figura 2.1, puede plantearse en la industria de bebidas. Supongamos que cada marca de cerveza se representa como un vector en un espacio definido por atributos, como *tradicional* versus *innovador* y *masivo* versus *premium*. En ese espacio, una lager clásica de gran consumo quedaría ubicada cerca de otras variedades tradicionales y de alcance masivo, mientras que una IPA artesanal o una edición limitada se situarían en la región asociada a lo *premium* e *innovador*. El sistema de recomendación aprovecha esta representación para calcular distancias o similaridades entre productos. Si un cliente suele elegir artículos situados en torno al cuadrante de *premium-tradicional*, el modelo infiere que probablemente muestre interés por otras marcas que ocupan posiciones cercanas en ese mismo espacio vectorial.

De esta manera, la proximidad entre vectores se convierte en un indicador de afinidad, que guía la generación de recomendaciones personalizadas.

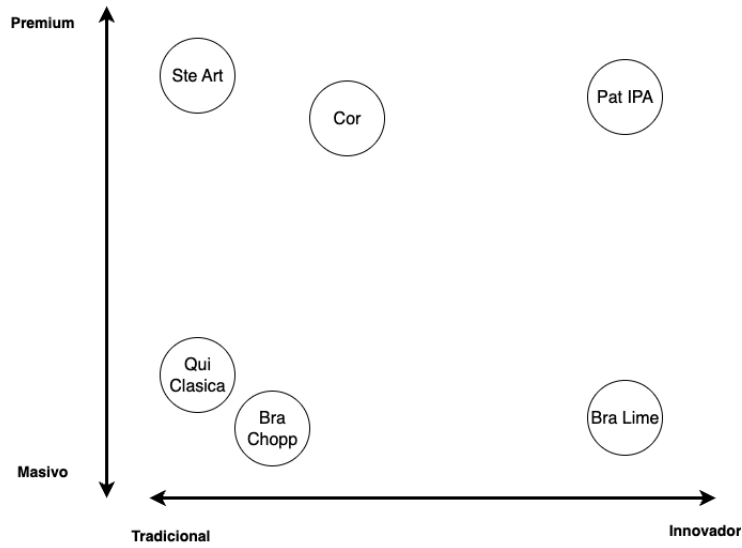


FIGURA 2.1. Ejemplo de representación de marcas de cerveza en un espacio de atributos.

2.1.2. Tipos de *feedback*

El tipo de información disponible para alimentar un sistema de recomendación también es determinante. Se distinguen dos formas principales de retroalimentación [7]. La retroalimentación explícita consiste en la valoración directa que realizan los clientes sobre los productos, como calificaciones numéricas, encuestas o reseñas. La retroalimentación implícita, en cambio, se infiere del comportamiento de los clientes, ya sea a través de sus compras, búsquedas o interacciones digitales. En el ámbito B2B, donde no es común que los clientes asignen calificaciones explícitas, predominan las señales implícitas, lo que plantea desafíos adicionales para la construcción de modelos precisos.

2.1.3. Filtrado colaborativo

El filtrado colaborativo se apoya en la hipótesis de que usuarios similares tienden a preferir ítems similares, lo que puede abordarse mediante enfoques *user-based* o *item-based* [8]. Su implementación moderna se basa en factorización matricial [9].

En la modalidad *user-based*, se recomienda a un cliente productos que fueron consumidos por otros con patrones de compra semejantes. En la modalidad *item-based*, se priorizan productos que suelen aparecer en conjunto en los historiales de distintos clientes.

El filtrado colaborativo suele implementarse mediante técnicas de factorización matricial. Dado un conjunto de m usuarios y n productos, se construye una matriz de interacciones $R \in \mathbb{R}^{m \times n}$, donde cada celda refleja el vínculo entre un cliente y un producto. El objetivo consiste en aproximar esta matriz como el producto de dos matrices de menor dimensión, como se puede observar en 2.1.

$$R \approx U \cdot V^T \quad (2.1)$$

donde $U \in \mathbb{R}^{m \times k}$ representa a los usuarios en un espacio latente de dimensión k , y $V \in \mathbb{R}^{n \times k}$ representa a los ítems en ese mismo espacio. La predicción de la afinidad del usuario i con el ítem j se calcula en 2.2 como el producto escalar entre los vectores latentes correspondientes.

$$\hat{r}_{ij} = u_i \cdot v_j^T \quad (2.2)$$

Este modelo permite capturar relaciones complejas entre clientes y productos a partir de información implícita, aunque presenta limitaciones frente al problema del arranque en frío, cuando no existe historial suficiente de interacciones.

2.1.4. Sistemas basados en contenido

Otro enfoque ampliamente utilizado es el de los sistemas basados en contenido, que centran la recomendación en las características de los productos y en el perfil de cada cliente [10]. En este caso, se representa a cada producto por un vector de atributos y se construye un perfil para cada cliente que refleja la importancia relativa de esos atributos en función de sus elecciones pasadas. La predicción de relevancia para recomendar un producto j a un cliente i puede expresarse de manera simplificada como en 2.3.

$$\hat{r}_{ij} = w_i \cdot x_j \quad (2.3)$$

donde x_j es el vector de atributos del producto y w_i representa el perfil del cliente. Este método permite recomendar productos nuevos o poco frecuentes siempre que exista información suficiente sobre sus atributos, lo que lo convierte en un complemento natural del filtrado colaborativo.

2.2. Fuentes de información

El funcionamiento de los sistemas de recomendación se apoya en diversas fuentes de información [5, 6] que, al combinarse, permiten construir una representación más completa de la relación entre usuarios y productos. Estas fuentes pueden clasificarse en tres grandes categorías: datos transaccionales, señales de interacción y atributos contextuales.

Los datos transaccionales reflejan las operaciones efectivamente realizadas, como compras, alquileres o reproducciones. Constituyen una evidencia directa de preferencia, ya que expresan decisiones concretas de los usuarios respecto a determinados productos o servicios.

Las señales de interacción incluyen registros de comportamiento que no necesariamente culminan en una transacción, pero que aportan información implícita de interés. Las señales implícitas, como visualizaciones o clics, resultan especialmente valiosas en contextos digitales [7, 11]. Ejemplos de este tipo de datos son las visualizaciones de fichas de producto, las búsquedas realizadas en una plataforma, las adiciones y eliminaciones en un carrito digital o las calificaciones otorgadas. Estas interacciones permiten identificar patrones de consideración más allá de la compra final.

Finalmente, los atributos contextuales corresponden a características adicionales tanto de los usuarios como de los productos. Del lado de los usuarios, se pueden incluir variables demográficas, geográficas o vinculadas al canal de consumo. Del lado de los productos, se consideran atributos como categoría, marca, segmento o características técnicas. Este conjunto de información enriquece la representación de afinidad, al capturar heterogeneidades que condicionan las recomendaciones.

De esta forma, la integración de datos transaccionales, señales de interacción y atributos contextuales constituye la base informativa sobre la cual se construyen los diferentes enfoques de recomendación. La disponibilidad y calidad de estas fuentes son determinantes para el desempeño de los modelos y para la capacidad de generar sugerencias precisas y relevantes.

2.3. Herramientas utilizadas

El desarrollo de este trabajo se apoyó en un conjunto de herramientas tecnológicas que facilitaron la gestión integral del ciclo de vida del sistema de recomendación. A continuación, se detallan las principales plataformas, bibliotecas y entornos empleados, junto con su función específica en el proceso.

2.3.1. Plataformas de procesamiento distribuido

El procesamiento y consolidación de grandes volúmenes de información se llevó a cabo en la plataforma Databricks [12], que integra un entorno colaborativo con un motor de cómputo distribuido basado en Apache Spark [13]. Esta herramienta permitió orquestar la ingestión de datos, ejecutar transformaciones a gran escala mediante PySpark y garantizar reproducibilidad en los flujos de trabajo. El uso de Databricks resultó fundamental para integrar múltiples fuentes y preparar los insumos que alimentaron las etapas de análisis y modelado.

2.3.2. Gestión del ciclo de vida de modelos

Para la gestión del ciclo de vida de los modelos se empleó MLflow [14], plataforma que facilita el registro de experimentos, parámetros, métricas y versiones de modelos. Esta herramienta permitió mantener trazabilidad entre las distintas ejecuciones, asegurar comparabilidad de resultados y almacenar los artefactos generados (modelos entrenados y estructuras derivadas). De este modo, se consolidó un repositorio ordenado que garantizó reproducibilidad y control en la experimentación.

2.3.3. bibliotecas de aprendizaje automático y profundo

En el desarrollo de los modelos se emplearon distintas bibliotecas que constituyen estándares en la comunidad científica y profesional. Se utilizó MLlib para implementar el filtrado colaborativo mediante factorización matricial con el algoritmo Alternating Least Squares (ALS) [15], mientras que LightFM [16] permitió construir un modelo híbrido que combina señales de interacción implícita con atributos de clientes y productos. Adicionalmente, se recurrió a PyTorch como entorno de deep learning para el diseño de arquitecturas neuronales capaces de capturar relaciones no lineales y complejas en los datos.

2.3.4. bibliotecas de visualización

Para el análisis visual y la generación de gráficos se utilizaron bibliotecas como Matplotlib [17] y Seaborn [18], que facilitaron la representación gráfica tanto de la información explorada como de los resultados obtenidos en las distintas fases del trabajo.

2.3.5. Control de versiones y colaboración

La organización y versionado del código se gestionaron mediante GitHub [19], que permitió estructurar los repositorios, registrar cambios de manera sistemática y facilitar la colaboración. El uso de esta plataforma aseguró orden en el desarrollo, trazabilidad de modificaciones y una integración eficiente de los distintos componentes del sistema.

2.3.6. Consideraciones finales

En conjunto, estas herramientas brindaron una infraestructura robusta para abordar todas las etapas del desarrollo del sistema de recomendación, desde la preparación de los datos hasta la evaluación y almacenamiento de modelos. Cabe destacar que la calidad de una implementación no depende únicamente del algoritmo utilizado, sino también de la solidez del entorno técnico que la respalda. El uso articulado de estas herramientas permitió asegurar la reproducibilidad de los resultados, la eficiencia en el manejo de datos, la trazabilidad de las decisiones y la escalabilidad del sistema desarrollado. En el contexto de una solución real, contar con esta base técnica resulta clave para garantizar tanto la calidad técnica como la posibilidad de evolución futura del sistema.

Capítulo 3

Diseño e implementación

Este capítulo aborda el proceso de construcción del sistema de recomendación, desde la concepción de la solución hasta su materialización en un entorno operativo. Se presentan las principales decisiones de diseño, el tratamiento de los datos y las técnicas empleadas para generar las recomendaciones, así como los lineamientos seguidos para asegurar que la propuesta resulte escalable, reproducible y alineada con los objetivos del negocio.

3.1. Diseño de solución

El sistema de recomendación se diseñó con el objetivo de estimar la afinidad entre clientes y productos en un entorno caracterizado por alta escala, heterogeneidad de perfiles y rotación constante del portafolio. El diseño de la solución se apoyó en una arquitectura en capas que permite integrar diversas fuentes de datos, transformarlas en estructuras analíticas consistentes, entrenar modelos capaces de capturar relaciones complejas y, finalmente, desplegar las recomendaciones en un flujo operativo estable y reproducible.

La primera capa corresponde a la ingestión de datos, instancia en la que se integran registros transaccionales, interacciones digitales y atributos contextuales. Los datos transaccionales reflejan las compras efectivas realizadas en distintos horizontes temporales, lo que aporta evidencia directa sobre las preferencias observadas. Las interacciones digitales, en cambio, ofrecen señales implícitas de interés a partir de búsquedas, visualizaciones de productos o modificaciones en el carrito. Finalmente, los atributos contextuales caracterizan tanto a los clientes, mediante variables asociadas a su canal de comercialización, localización o tamaño, como a los productos, a partir de propiedades como marca, categoría o segmento.

La segunda capa se orienta a la preparación de los datos. En esta etapa se construyen las matrices de interacciones cliente–producto y se generan representaciones temporales que permiten capturar la dinámica de la demanda. Asimismo, se aplican técnicas de tratamiento de valores faltantes y de codificación de atributos categóricos, con el fin de asegurar consistencia y compatibilidad entre las distintas fuentes. El resultado de este proceso es un conjunto de estructuras homogéneas que sientan las bases para la etapa de modelado.

El modelado constituye la tercera capa de la arquitectura. En este punto se combinan distintos enfoques con el fin de maximizar la capacidad predictiva y superar

las limitaciones de cada técnica individual. El filtrado colaborativo implícito, implementado a través de factorización matricial con el método ALS, permite capturar patrones latentes a partir de historiales de compra extensos. Los modelos basados en contenido complementan este enfoque al aprovechar descripciones de clientes y productos, y ofrecen una alternativa frente al problema del arranque en frío. Adicionalmente, se exploran modelos híbridos y de aprendizaje profundo capaces de integrar simultáneamente señales transaccionales y digitales, y de modelar relaciones no lineales entre las variables.

Finalmente, la capa de despliegue asegura la integración del motor de recomendación en el ecosistema tecnológico de la empresa. El pipeline resultante genera listas de productos priorizados para cada cliente, incorpora mecanismos de versionado y monitoreo de modelos, y permite evaluar su desempeño en forma continua. De este modo, la solución se diseñó no solo para alcanzar precisión en la generación de recomendaciones, sino también para garantizar escalabilidad, reproducibilidad y adaptabilidad frente a la evolución del portafolio y a los cambios en los objetivos estratégicos.

La arquitectura de la solución se representa en la figura 3.1. Allí se observa el flujo general del sistema, desde la integración de datos hasta la generación de recomendaciones. El diagrama sintetiza los módulos principales y sus interacciones, y ofrece una visión global que facilita comprender cómo se organiza el motor de afinidad.

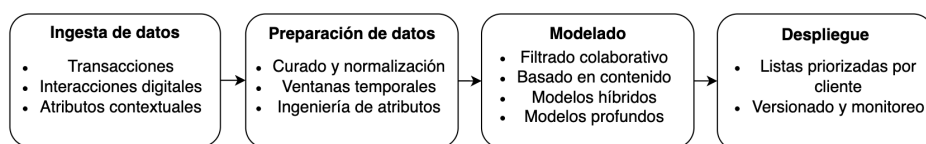


FIGURA 3.1. Arquitectura de alto nivel del sistema de recomendación.

3.2. Preparación de los datos

La preparación de los datos se centró en transformar las distintas fuentes disponibles en insumos consistentes y comparables para el entrenamiento de los modelos de recomendación. El proceso partió de tres conjuntos principales de información: registros transaccionales, eventos digitales generados en la aplicación y atributos contextuales de clientes y productos. La combinación de estas fuentes permitió construir una representación integral de la relación cliente–producto, en la que se entrelazan tanto preferencias explícitas como señales implícitas de interés.

Los registros transaccionales se encontraban a nivel de operación individual, con granularidad diaria y asociados a identificadores de cliente, producto, cantidad y monto. Este tipo de información es particularmente relevante en entornos de consumo masivo B2B, donde suele observarse una fuerte concentración de las compras en un subconjunto reducido de productos y clientes. De manera preliminar, se advierte un patrón cercano a la regla de Pareto [20], un pequeño grupo de marcas concentra la mayor parte del volumen mientras que muchos otros artículos registran ventas esporádicas. Esta característica, común en el sector, anticipa la necesidad de abordar los sesgos hacia productos de alta rotación en etapas posteriores del análisis.

Los eventos digitales, por su parte, se encontraban a nivel de interacción, con registros de búsquedas, visualizaciones, adiciones y remociones en el carrito, y clics en promociones. Estos datos permiten capturar señales tempranas de interés que no siempre se traducen en transacciones. No obstante, su naturaleza los hace sensibles al ruido: interacciones aisladas, comportamientos exploratorios o promociones masivas pueden generar señales que no representan un patrón estable. En consecuencia, aunque aportan cobertura y diversidad, requieren una interpretación cuidadosa para evitar que se sobreestimen comportamientos circunstanciales.

Finalmente, los atributos contextuales ofrecieron información complementaria a nivel de cliente y producto. En el caso de los clientes, la granularidad fue de punto de venta, con variables que reflejan diferencias de canal, localización o tamaño, factores que suelen condicionar significativamente las decisiones de compra. En el caso de los productos, los atributos de marca, segmento, envase o rango de precio permiten distinguir entre artículos de consumo masivo y aquellos de carácter más selectivo, lo que introduce heterogeneidad que no siempre se refleja en los registros transaccionales.

La integración de estas fuentes en un repositorio unificado aseguró la compatibilidad de identificadores y la alineación temporal de los registros, lo que habilitó su uso conjunto en el análisis. Las primeras observaciones sugieren un escenario donde coexisten concentración de la demanda en pocos productos y clientes, señales digitales útiles pero ruidosas y una marcada diversidad contextual. Estos fenómenos, característicos del consumo masivo B2B, definen los ejes que serán explorados en mayor detalle durante el análisis exploratorio y que posteriormente condicionarán las decisiones de ingeniería de atributos y modelado.

3.3. Análisis exploratorio de los datos

El análisis exploratorio constituye una etapa fundamental para comprender la estructura y los patrones subyacentes en la información disponible antes de su utilización en modelos de recomendación. Su objetivo es identificar distribuciones, tendencias y relaciones entre variables que permitan caracterizar el comportamiento de los clientes y del portafolio de productos, así como anticipar posibles limitaciones o sesgos que afecten el desempeño de los algoritmos.

En esta sección se examinan distintas dimensiones de los datos, e incluye la concentración de clientes y productos, la diversidad de los portafolios de compra, las correlaciones entre variables transaccionales y digitales, y la presencia de sesgos asociados a la popularidad. Este análisis preliminar no solo proporciona una visión descriptiva del conjunto de datos, sino que también orienta decisiones posteriores de ingeniería de atributos y diseño de modelos, al revelar qué señales resultan más informativas y qué fenómenos requieren un tratamiento específico.

3.3.1. Curvas de concentración de clientes y productos

El análisis de concentración constituye un paso clave para comprender la distribución del consumo en entornos de negocio masivo. La figura 3.2 muestra la curva de concentración de productos, donde se observa que un reducido conjunto concentra la mayor parte del volumen total. En particular, el 20 % de los productos explica cerca del 90 % de las ventas acumuladas, mientras que el resto

conforma una extensa cola larga con niveles de rotación significativamente menores. Este comportamiento coincide con la ley de Pareto o principio 80/20 [20], ampliamente documentado en mercados de consumo masivo, donde la dinámica competitiva se organiza en torno a un pequeño núcleo de artículos de alta popularidad y una mayoría de baja incidencia [21].

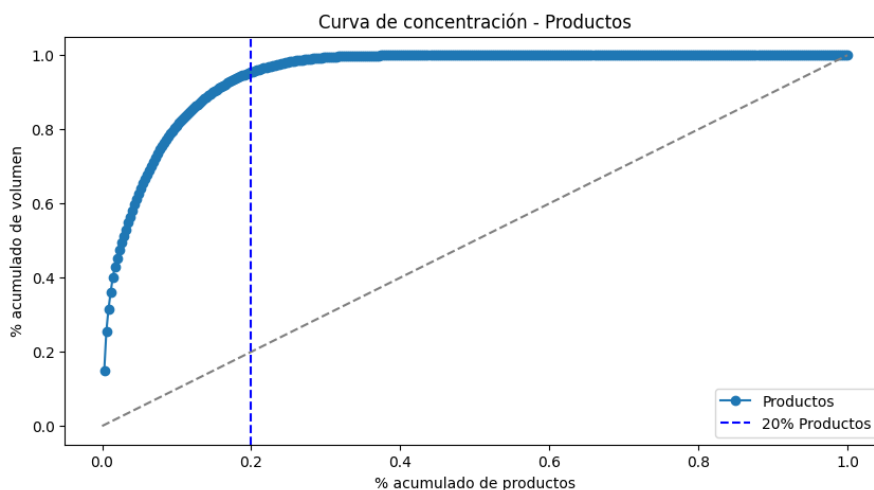


FIGURA 3.2. Concentración de productos en el portafolio.

De manera análoga, la figura 3.3 refleja la concentración del consumo en la base de clientes. Los resultados indican que cerca del 20 % de los puntos de venta generen alrededor del 80 % del volumen total, lo que pone de manifiesto la existencia de clientes estratégicos que concentran gran parte de la demanda. Esta distribución desigual plantea desafíos relevantes para el diseño de sistemas de recomendación, ya que las señales provenientes de clientes de alto volumen tienden a dominar los modelos, lo que genera sesgos hacia productos y comportamientos mayoritarios.

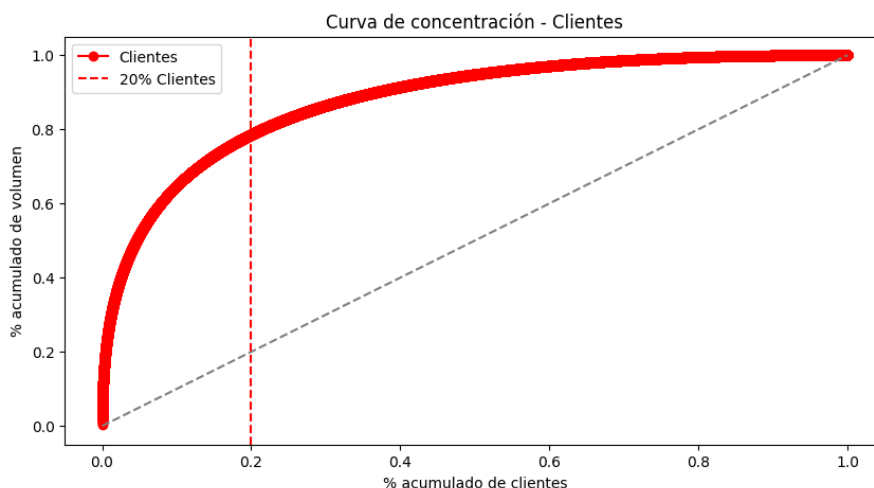


FIGURA 3.3. Concentración de clientes.

La evidencia empírica confirma así que tanto el portafolio de productos como la base de clientes presentan fuertes patrones de concentración. En consecuencia, un motor de recomendación que busque maximizar su impacto no solo debe capturar la afinidad entre clientes y productos más relevantes, sino también considerar

mecanismos que favorezcan la diversidad y la exploración de la cola larga. Esta perspectiva resulta fundamental para equilibrar la explotación de los artículos de mayor rotación con la exposición de productos menos populares, lo que alinea los objetivos de negocio con la mejora de la experiencia del cliente.

3.3.2. Patrones de diversidad en el portafolio

El análisis de la diversidad en el portafolio de productos por cliente permite comprender la amplitud y heterogeneidad de los hábitos de consumo. La figura 3.4 muestra la distribución del número de productos distintos adquiridos por cliente en un mes. Los resultados evidencian que la mayoría de los puntos de venta concentra su demanda en un conjunto reducido de referencias, mientras que un número menor incorpora una mayor amplitud de marcas y presentaciones. Esta asimetría confirma la coexistencia de clientes de bajo rango de exploración con otros de portafolio más diversificado.

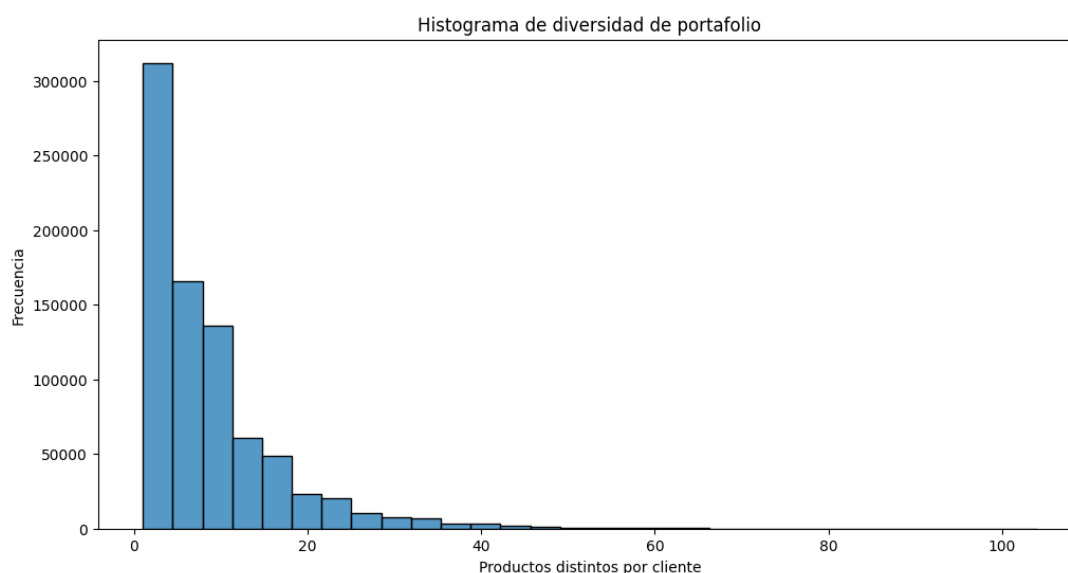


FIGURA 3.4. Histograma de diversidad de portafolio: número de productos distintos por cliente.

Las diferencias se acentúan al segmentar por canal comercial, como se puede apreciar en la figura 3.5. En este caso, se observa que los autoservicios tienden a manejar un surtido más amplio de productos en comparación con kioscos y tiendas tradicionales, lo que refleja el rol que cada formato cumple dentro de la red de distribución. Este hallazgo es consistente con la literatura en consumo masivo, que indica que la variedad de portafolio suele estar asociada a factores estructurales como el tamaño del punto de venta y la frecuencia de reposición [22].

La figura 3.6 ilustra el fenómeno de concentración extrema en la demanda, donde unos pocos productos acumulan la mayoría de los pedidos mientras que la gran mayoría registra volúmenes marginales. Para representar este patrón se utiliza un *log-log plot*, en el cual tanto el ranking de los productos como su número total de pedidos se expresan en escala logarítmica. Esta transformación permite visualizar con mayor claridad distribuciones de tipo cola larga, que en escalas lineales suelen quedar ocultas por la presencia de artículos extremadamente populares.

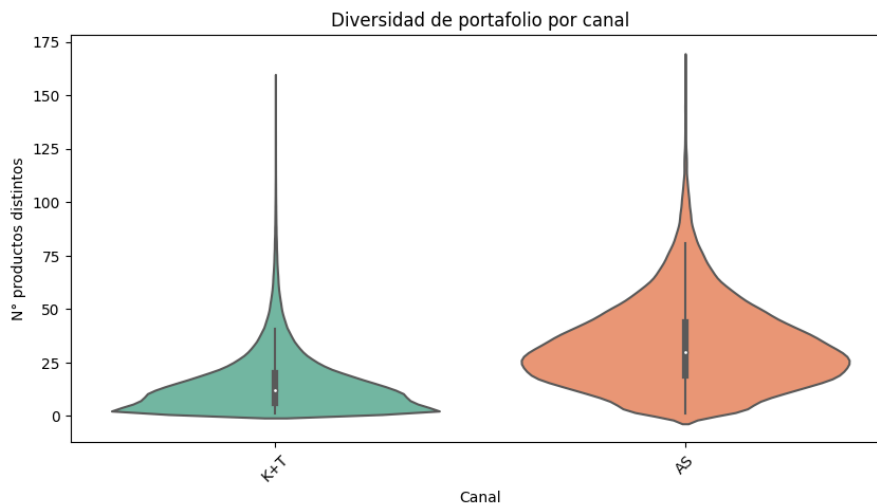


FIGURA 3.5. Diversidad de portafolio segmentada por canal comercial.

El gráfico muestra una pendiente decreciente que confirma la existencia de este comportamiento: un reducido conjunto de productos concentra un volumen muy elevado, mientras que el resto se distribuye en la larga cola de baja rotación.

Este patrón no solo refuerza la evidencia presentada en las curvas de concentración, sino que además resalta un sesgo estructural que enfrenta cualquier sistema de recomendación en entornos de consumo masivo. Al entrenarse sobre datos históricos, los modelos tienden de manera natural a privilegiar los productos más populares, lo que reproduce el sesgo de popularidad y reduce la diversidad de las sugerencias. Este fenómeno señala la tensión entre explotación de productos estrella y exploración de la cola larga [23]. En este contexto, el desafío consiste en diseñar mecanismos que permitan balancear ambos extremos, de modo que se garantice relevancia sin sacrificar diversidad ni cobertura.

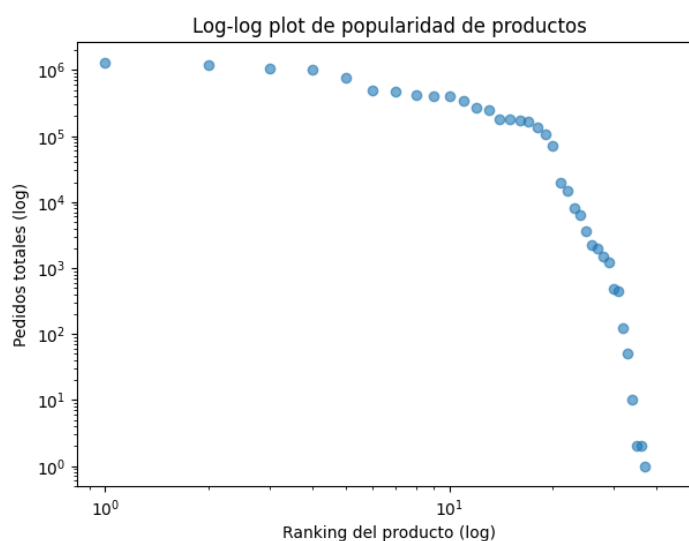


FIGURA 3.6. *Log-log plot* de la popularidad de productos.

El análisis de co-ocurrencia entre los productos más relevantes, presentado en la figura 3.7, revela patrones de complementariedad en la demanda. Determinadas

marcas y presentaciones tienden a aparecer de manera conjunta en los carritos de compra, lo que sugiere asociaciones naturales que pueden ser aprovechadas por un motor de recomendación. Estos resultados refuerzan la importancia de capturar no solo la popularidad individual de cada producto, sino también las relaciones de afinidad que emergen a nivel de portafolio.

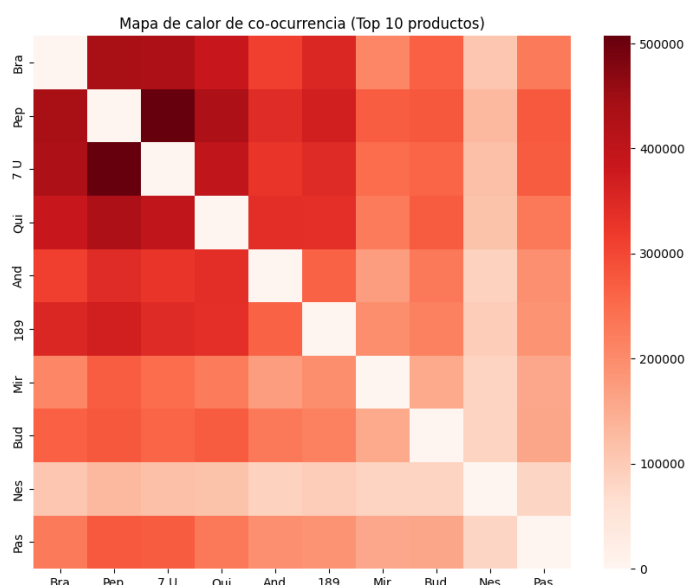


FIGURA 3.7. Mapa de calor de co-ocurrencia entre los 10 productos más relevantes.

3.3.3. Correlaciones entre variables transaccionales y digitales

El análisis de correlaciones busca identificar hasta qué punto las señales digitales anticipan comportamientos de compra y, en consecuencia, evaluar su potencial como insumos predictivos. Con el fin de evaluar la relación entre interacciones digitales y transacciones, se construyó una matriz de correlación entre las principales variables del conjunto de datos, observada en la figura 3.8.

Los resultados muestran una correlación elevada entre `ORDERED` y `BUYER` ($r = 0,70$), coherente con el hecho de que ambas variables reflejan distintos aspectos de la misma dimensión de compra. En contraste, las correlaciones de las señales digitales con las variables de compra resultan positivas pero de menor magnitud: `CARD_VIEWED` y `DETAILS_PAGE_VIEWED` muestran coeficientes bajos, lo que indica que la exposición y exploración de productos acompaña el proceso de compra, aunque no lo determina. La variable `REMOVED` presenta la relación más débil, lo que sugiere que los eventos de descarte contienen información ruidosa y limitada respecto de la propensión a comprar.

La correlación contemporánea entre interacciones digitales y compras confirma que las transacciones pasadas siguen siendo el principal indicador de comportamiento, mientras que las señales digitales aportan evidencia complementaria que, si bien débil de manera aislada, resulta relevante al integrarse en un modelo híbrido.

Con el fin de explorar la capacidad predictiva de estas variables, se calculó la correlación de cada una con la compra del mismo cliente-producto en el mes siguiente. Los resultados, en la figura 3.9, muestran que las transacciones pasadas

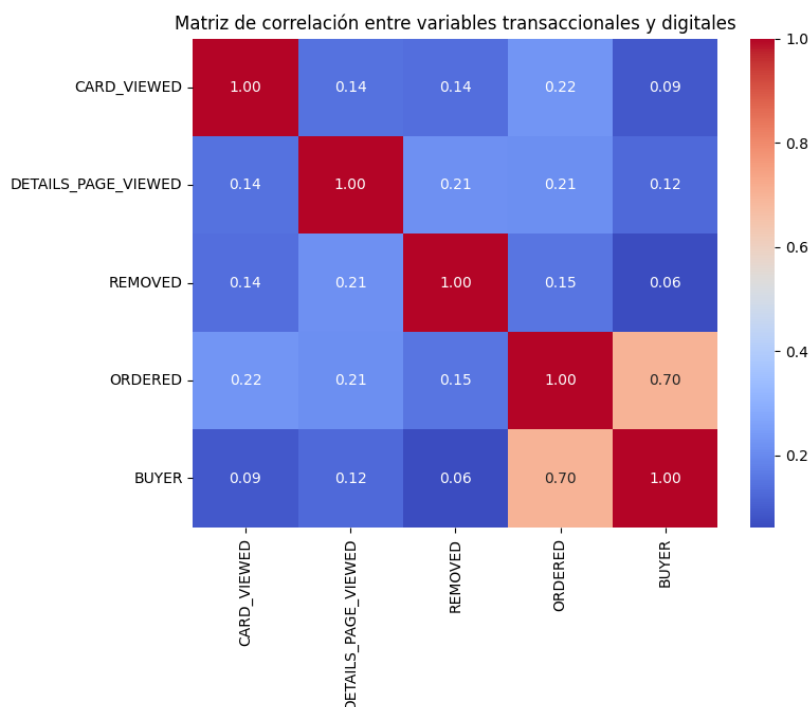


FIGURA 3.8. Matriz de correlación entre variables transaccionales y digitales.

(BUYER, ORDERED) son los predictores más fuertes, aunque las señales digitales también aportan información incremental. En particular, la variable `CARD_VIEWED` presenta un coeficiente relevante, lo que respalda la hipótesis de que la exposición reiterada a un producto incrementa la probabilidad de recompra.

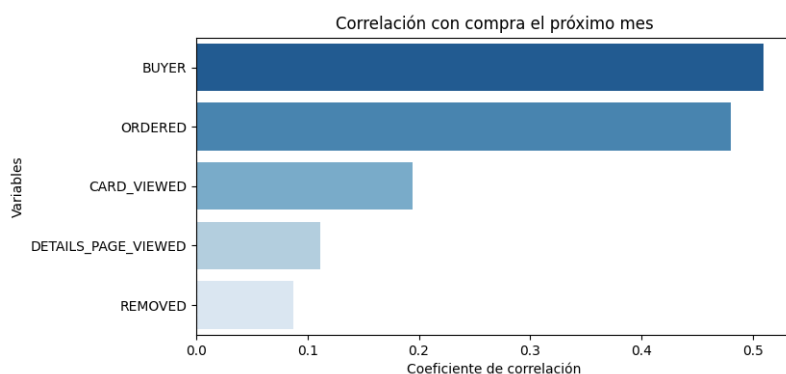


FIGURA 3.9. Correlación de variables con la compra en el mes siguiente.

Finalmente, se evaluó la tasa de recompra según la combinación de señales observadas en meses previos, en la tabla 3.1. Los clientes que registran tanto interacción como transacción presentan la mayor tasa de recompra (58,8%), seguidos por aquellos con solo órdenes (44,1%). En contraste, quienes solo exhiben interacciones digitales alcanzan un nivel considerablemente menor (17,6%), incluso por debajo del grupo sin ningún registro previo (28,3%). Este resultado sugiere que las interacciones aisladas no constituyen un predictor confiable de recompra,

sino que tienden a reflejar un interés superficial que rara vez se traduce en pedidos. En cambio, la combinación de transacciones previas con señales digitales se confirma como el escenario de mayor poder explicativo, ya que aprovecha la solidez de la evidencia transaccional y, al mismo tiempo, permite mejorar la capacidad de anticipar comportamientos futuros en casos donde no existen registros abundantes de compra.

TABLA 3.1. Tasa de recompra según la combinación de señales previas.

Grupo	Tasa de recompra
Interacción y orden	58.79 %
Solo orden	44.05 %
Ninguno	28.30 %
Solo interacción	17.64 %

3.3.4. Observaciones preliminares del análisis exploratorio

El análisis exploratorio permitió identificar una serie de patrones que resultan fundamentales para orientar el diseño del motor de recomendación. En primer lugar, se confirmó que tanto el portafolio de productos como la base de clientes presentan fuertes niveles de concentración: un reducido conjunto explica la mayor parte del volumen, mientras que la mayoría se distribuye en una extensa cola larga de baja rotación. Este comportamiento introduce un sesgo hacia popularidad que los modelos deben manejar para no sacrificar diversidad [21].

En segundo lugar, se observó que la diversidad en los portafolios de compra varía según el tipo de cliente, con autoservicios que incorporan un surtido más amplio en comparación con kioscos y tiendas tradicionales. Además, los análisis de co-ocurrencia revelaron asociaciones frecuentes entre ciertos productos, lo que sugiere la existencia de complementariedades que pueden ser aprovechadas en la generación de recomendaciones.

En tercer lugar, el estudio de correlaciones entre variables digitales y transaccionales mostró que, si bien las transacciones pasadas constituyen el predictor más sólido de comportamiento, las señales digitales aportan información incremental y se vuelven especialmente relevantes en escenarios de arranque en frío. La evaluación de tasas de recompra confirmó que la combinación de interacciones y compras pasadas es la fuente más robusta de predicción, mientras que las interacciones aisladas presentan un valor explicativo limitado.

En conjunto, estos hallazgos proporcionan una primera validación de la hipótesis central: la integración de señales transaccionales y digitales, complementadas con atributos contextuales, resulta clave para capturar la heterogeneidad del consumo y diseñar un motor de recomendación capaz de balancear precisión, diversidad y cobertura.

3.4. Ingeniería de atributos

La ingeniería de atributos busca transformar los datos en representaciones que los algoritmos de recomendación puedan utilizar de manera efectiva. Más que

un proceso técnico de limpieza o transformación, implica decidir cómo capturar las relaciones entre clientes y productos para que el modelo cuente con señales relevantes al momento de aprender.

En un entorno B2B de consumo masivo, estas decisiones resultan especialmente importantes. La concentración de la demanda en pocos productos, la rotación frecuente del portafolio y la existencia de clientes sin historial son factores que obligan a diseñar atributos que reflejen esa complejidad. Para ello se definieron distintas ventanas temporales de observación, se combinaron transacciones con interacciones digitales y se incorporaron características contextuales que enriquecen la representación final.

También se aplicaron técnicas de agregación, codificación y normalización para equilibrar las variables y evitar que los productos más populares concentren toda la señal. El resultado es una matriz cliente-producto adecuada para algoritmos como ALS o LightFM, capaz de capturar patrones de afinidad y ofrecer recomendaciones más precisas.

3.4.1. Ventanas temporales de datos

Uno de los primeros aspectos a definir en la ingeniería de atributos es la ventana temporal sobre la que se mide el comportamiento de los clientes. La dinámica de compra en el canal B2B de consumo masivo no es homogénea: algunos puntos de venta realizan pedidos de manera semanal, mientras que otros concentran sus compras una vez al mes o incluso con menor frecuencia. Capturar estas diferencias requiere observar la información en distintos horizontes.

Para este trabajo se construyeron indicadores en tres escalas: corto plazo definido como un mes, mediano plazo correspondiente a tres meses y largo plazo equivalente a seis meses. La ventana de un mes refleja comportamientos inmediatos y ayuda a identificar tendencias recientes, como la adopción de un producto nuevo o la respuesta a una promoción puntual. La ventana de tres meses suaviza la volatilidad y permite reconocer patrones de consumo más estables. Finalmente, la de seis meses ofrece una visión de mayor persistencia, útil para productos de baja rotación o con fuerte estacionalidad.

El uso combinado de estas ventanas facilita que el modelo integre señales de distinto alcance temporal [24, 25], lo que equilibra la sensibilidad a cambios recientes con la estabilidad de patrones históricos. Así, un cliente que dejó de comprar un producto en el último mes pero que lo consumía de manera regular en períodos anteriores puede ser identificado de forma diferente a uno que nunca lo incorporó a su portafolio. De esta manera, se evita que las recomendaciones se basen únicamente en la foto más reciente y se logra un balance entre actualidad y consistencia histórica.

3.4.2. Agregación de características

Una vez definidas las variables de interés, fue necesario condensar la información en indicadores que representen de manera clara el comportamiento de cada cliente frente a cada producto. En las transacciones se calcularon medidas como el número total de órdenes, el volumen acumulado y la participación relativa de cada producto en el portafolio del cliente [6].

Las interacciones digitales se resumieron en recuentos de eventos significativos, tales como visualizaciones de producto, adiciones o remociones en el carrito y respuestas a promociones. Este tipo de métricas permiten obtener una señal cuantitativa de interés que complementa la evidencia proveniente de las compras efectivas [5].

Además, se incorporaron características contextuales tanto del lado del cliente como del producto. Para los clientes se consideraron variables como canal comercial, localización geográfica y tamaño del punto de venta. Para los productos, atributos como marca, segmento, envase y rango de precio [26]. Estas propiedades enriquecen la caracterización de cada par cliente–producto y permiten capturar heterogeneidades relevantes para el sistema de recomendación.

El resultado es un conjunto de variables agregadas que sintetizan de manera equilibrada la información transaccional, digital y contextual, lo que conforma un insumo más robusto y manejable para la construcción de la matriz cliente–producto.

3.4.3. Codificación de variables categóricas

El conjunto de atributos contextuales incluyó variables de naturaleza categórica, tales como canal comercial, región geográfica, segmento de producto, marca o tipo de envase. Para que estas variables pudieran ser utilizadas en los algoritmos de recomendación fue necesario transformarlas en representaciones numéricas que preservaran su información.

En casos de variables con gran número de categorías, como la marca o el punto de venta, se aplicaron criterios de reducción para evitar una explosión dimensional. Las categorías de baja frecuencia se reagruparon en clases residuales, lo que garantizó que la codificación reflejara solo valores con suficiente representación en los datos. Este enfoque reduce la dispersión y mejora la estabilidad del entrenamiento, al tiempo que mantiene la capacidad de capturar patrones relevantes.

La estrategia aplicada fue la indexación de categorías [5], que asigna a cada valor un identificador numérico único dentro de su variable. De esta forma, un producto perteneciente a un determinado segmento o una venta registrada en un canal específico quedan representados de manera consistente.

Se obtuvo un conjunto de variables categóricas estandarizadas y codificadas, listas para integrarse a la matriz cliente–producto junto con las métricas transaccionales y digitales. De esta manera, la representación final no solo incorpora información sobre la intensidad de las interacciones, sino también sobre el contexto en que estas ocurren.

3.4.4. Tratamiento de valores atípicos

El análisis preliminar de los datos reveló la presencia de valores atípicos tanto en las variables transaccionales como en las digitales. Este fenómeno es esperable en un entorno B2B de consumo masivo, donde conviven puntos de venta de muy distinto tamaño y productos con dinámicas de demanda altamente desbalanceadas. Sin un tratamiento adecuado, estos registros extremos podían distorsionar las métricas agregadas y sesgar el aprendizaje de los modelos de recomendación [27].

En el caso de las transacciones, se observaron clientes con volúmenes de compra extraordinariamente altos en comparación con el resto, muchas veces asociados a distribuidores mayoristas o a operaciones excepcionales. Para mitigar este efecto, se aplicaron reglas de truncamiento que limitaron las observaciones a rangos definidos por percentiles superiores e inferiores, lo que preservó la distribución general sin permitir que unos pocos casos dominaran la representación.

En las interacciones digitales, los valores atípicos se manifestaron principalmente en clientes con una actividad anómala, caracterizada por cientos de visualizaciones o clics en periodos muy cortos. Estos registros fueron depurados mediante filtros basados en umbrales máximos por tipo de evento y cliente, lo que aseguró que las métricas reflejaran comportamientos consistentes con el resto de la población.

En el caso de los atributos contextuales, se verificó la existencia de categorías con ocurrencias aisladas o inconsistentes, que fueron unificadas o eliminadas según correspondiera. Este procedimiento redujo la dispersión y permitió consolidar categorías con suficiente representación estadística.

Con estas medidas se logró un conjunto de datos más estable y representativo, en el que las señales predominantes no se ven opacadas por comportamientos excepcionales.

3.4.5. Normalización y escalado

Las variables construidas a partir de transacciones e interacciones digitales presentaban rangos y magnitudes muy heterogéneos. Mientras que algunas métricas se expresaban en unidades absolutas de volumen o número de órdenes, otras correspondían a recuentos de eventos digitales con una escala mucho menor. Si estas diferencias no se trataban de manera adecuada, existía el riesgo de que los modelos de recomendación priorizaran de forma desproporcionada a las variables de mayor rango numérico, lo que hacía que se perdiera capacidad para capturar señales más sutiles.

Para abordar este problema se aplicaron procedimientos de normalización y escalado [26]. En primer lugar, se transformaron los recuentos absolutos en tasas relativas, como la proporción de un producto dentro del portafolio de cada cliente o el peso de una categoría de interacción sobre el total de eventos registrados. Este ajuste permitió comparar clientes con distintos niveles de actividad sin sesgos derivados de su tamaño.

En segundo lugar, las variables continuas se escalaron a rangos comparables mediante técnicas de estandarización, de forma que cada atributo contribuyera de manera equilibrada al entrenamiento. En particular, se utilizaron transformaciones que preservan la distribución relativa de las observaciones, lo que evita la pérdida de información sobre diferencias significativas entre clientes o productos.

De esta forma se obtuvo un conjunto de características homogéneas y comparables, con un balance adecuado entre variables. Esto permitió que los algoritmos de recomendación incorporaran de manera conjunta tanto las señales transaccionales como las digitales y contextuales, sin que ninguna dominara de forma artificial sobre el resto.

3.4.6. Representación final de cliente-producto

La aplicación conjunta de estas etapas permitió transformar la heterogeneidad de las fuentes originales en un espacio de características consistente y utilizable por los modelos de recomendación. Cada par cliente-producto quedó representado por un vector que combina tres dimensiones complementarias: indicadores transaccionales que capturan la intensidad y estabilidad de las compras, métricas digitales que reflejan señales de interés implícito y atributos contextuales que sitúan la relación dentro de un marco comercial más amplio.

Las ventanas temporales aportaron una visión multinivel de la dinámica de consumo, las agregaciones resumieron la información en métricas robustas, el tratamiento de valores atípicos redujo la influencia de casos extremos, y la codificación de variables categóricas permitió integrar atributos cualitativos en un formato numérico. Finalmente, la normalización y el escalado homogenizaron magnitudes, lo que garantizó que ninguna señal predomine artificialmente sobre las demás.

De manera simplificada, la representación final de la relación entre el cliente i y el producto j puede expresarse como en 3.1 donde $T_{ij}^{(k)}$ corresponde a los indicadores transaccionales calculados en ventanas de k meses, $D_{ij}^{(k)}$ a los indicadores digitales en los mismos horizontes, y C_i, C_j a los atributos contextuales del cliente y del producto respectivamente.

$$X_{ij} = [T_{ij}^{(1)}, T_{ij}^{(3)}, T_{ij}^{(6)}, D_{ij}^{(1)}, D_{ij}^{(3)}, D_{ij}^{(6)}, C_i, C_j] \quad (3.1)$$

Así se conforma una matriz cliente-producto enriquecida, en la que cada celda refleja no solo la existencia de una interacción, sino también su intensidad, contexto y evolución temporal. Esta representación constituye la base sobre la cual se entrenarán los modelos de recomendación presentados en la sección siguiente.

3.5. Desarrollo de modelos

El desarrollo de los modelos constituye la fase central del sistema de recomendación, donde la matriz cliente-producto construida en etapas previas se transforma en un mecanismo capaz de estimar la afinidad entre ambos. El objetivo es asignar a cada combinación posible un puntaje continuo que refleje la probabilidad relativa de interés, permitiendo ordenar los productos según su relevancia esperada para cada cliente.

Para abordar este desafío se exploraron distintos enfoques de modelado, que combinan estrategias colaborativas, basadas en contenido y de aprendizaje profundo. En primer lugar, se implementó un modelo de filtrado colaborativo mediante el algoritmo *Alternating Least Squares* (ALS) [15, 7], que aprende representaciones latentes de clientes y productos a partir de los patrones de interacción observados. En segundo lugar, se desarrolló un modelo híbrido con *LightFM* [16], capaz de integrar señales colaborativas con atributos explícitos de clientes y productos, mitigando así el problema del arranque en frío.

Complementariamente, se incorporaron dos variantes basadas en redes neuronales. La primera aprende representaciones de clientes y productos a partir de sus atributos mediante arquitecturas de *embeddings* [28], que pueden combinarse con ALS en un esquema de ensamble. La segunda corresponde al enfoque de

Neural Collaborative Filtering (NCF) [29], que reemplaza la combinación lineal de embeddings por un modelo neuronal capaz de capturar relaciones no lineales de afinidad.

En conjunto, estos modelos representan una progresión desde métodos clásicos hacia aproximaciones más flexibles y expresivas. Las siguientes subsecciones describen los fundamentos, la formulación y las particularidades de cada uno, sentando las bases para la comparación de desempeño y costos que se desarrolla en el capítulo siguiente.

3.5.1. Filtrado colaborativo con ALS

El primer modelo desarrollado fue un esquema de filtrado colaborativo mediante factorización matricial con ALS para *feedback* implícito [15]. El objetivo fue representar a clientes y productos en un espacio latente de menor dimensión, en el que la afinidad estimada entre ambos pudiera expresarse como una medida de similitud entre sus vectores de representación. Para ello se utilizó la matriz cliente–producto construida en la etapa de ingeniería de atributos, donde cada celda sintetiza la intensidad de interacción entre ambos a partir de señales transaccionales y digitales.

El modelo se configuró para trabajar con datos de *feedback* implícito, considerando que la ausencia de interacción no implica una preferencia negativa, sino simplemente falta de evidencia. Este enfoque resulta especialmente adecuado en contextos donde las señales de interés se derivan del comportamiento observado y no de valoraciones explícitas. El entrenamiento se realizó sobre las interacciones agregadas en una ventana de seis meses, lo que permitió capturar patrones de compra más estables y reducir el impacto de la estacionalidad.

Durante el entrenamiento, el algoritmo alterna la optimización de las representaciones latentes de clientes y productos, ponderando cada observación según el nivel de confianza asignado a la interacción. De esta forma, las señales más frecuentes o consistentes tienen mayor influencia en el aprendizaje, mientras que las observaciones esporádicas o débiles aportan con menor peso. El proceso se repite hasta alcanzar la convergencia, resultando en dos conjuntos de vectores que capturan las preferencias subyacentes en los datos.

Para la calibración del modelo se realizó un proceso de *hyperparameter tuning* orientado a maximizar el área bajo la curva de característica operativa del receptor (ROC-AUC) en un conjunto de validación temporal. Se evaluaron combinaciones de dimensión latente, nivel de regularización, número de iteraciones y parámetro de confianza (`alpha`), buscando un equilibrio entre capacidad predictiva y eficiencia computacional. La validación se efectuó en ventanas móviles, entrenando sobre los seis meses previos y evaluando sobre el mes siguiente, lo que permitió analizar la estabilidad del desempeño y minimizar el riesgo de sobreajuste.

Una vez finalizado el entrenamiento, se calcularon los puntajes de afinidad entre cada cliente y producto y se generaron listas *Top-K* personalizadas, ordenando los productos de acuerdo con su relevancia estimada. En los casos sin historial previo, como clientes nuevos o productos recientemente incorporados, se aplicaron estrategias de respaldo basadas en la popularidad dentro de cada categoría y

en el comportamiento promedio de clientes con perfiles similares. Estas estrategias aseguraron la cobertura total del sistema, garantizando que todos los clientes recibieran recomendaciones coherentes y consistentes con su segmento.

3.5.2. Modelo híbrido con LightFM

El segundo modelo desarrollado fue un enfoque híbrido implementado con la biblioteca `LightFM` [16], que combina las ventajas del filtrado colaborativo y de los sistemas basados en contenido dentro de un mismo marco. A diferencia del ALS, que se apoya únicamente en la información implícita proveniente de las interacciones históricas, este modelo incorpora además atributos explícitos de clientes y productos, permitiendo generar recomendaciones incluso en escenarios con historial limitado o inexistente.

El principio del modelo consiste en representar tanto a los clientes como a los productos mediante vectores que combinan dos fuentes de información: las interacciones observadas y los atributos que los describen. Estos vectores se entrenan conjuntamente para maximizar la probabilidad de que los pares cliente–producto observados obtengan un puntaje de afinidad mayor que los no observados. De esta manera, el modelo aprende a identificar relaciones latentes entre los rasgos de los clientes y las características de los productos, integrando en un mismo espacio de representación señales colaborativas y de contenido.

En la implementación del trabajo, los atributos de producto incluyeron variables como marca, segmento, envase y rango de precio, mientras que del lado de los clientes se consideraron características como canal comercial y región geográfica. Estas variables fueron codificadas e integradas al modelo como características adicionales, lo que permitió ampliar la información disponible para los casos con baja densidad transaccional.

El modelo se entrenó bajo un esquema de *feedback* implícito utilizando una función de pérdida logística, que penaliza las predicciones incorrectas en función de la distancia entre los puntajes observados y los esperados. Se realizaron pruebas adicionales con la función de pérdida *Bayesian Personalized Ranking* (BPR), aunque se observó una convergencia más estable con la opción logística, manteniendo un desempeño comparable en métricas de ranking.

Los hiperparámetros principales, dimensión latente, tasa de aprendizaje y número de épocas, se ajustaron mediante búsqueda bayesiana, utilizando validación temporal con ventanas deslizantes. La evaluación se centró en maximizar el área bajo la curva ROC-AUC.

Finalmente, el modelo se aplicó sobre la misma matriz cliente–producto, complementada con los atributos adicionales. Los puntajes generados permitieron ordenar los productos según su relevancia esperada para cada cliente, obteniendo listas *Top-K* comparables en estructura y formato a las del modelo anterior. Este enfoque resultó particularmente valioso para mitigar los efectos del arranque en frío y para mejorar la diversidad de las recomendaciones en categorías con baja frecuencia de compra.

3.5.3. Modelo de embeddings neuronales

El tercer enfoque desarrollado se basó en el aprendizaje de representaciones densas de clientes y productos mediante redes neuronales. Este modelo tuvo como

objetivo capturar relaciones no lineales en los patrones de comportamiento y complementar el enfoque de factorización matricial, que asume una estructura lineal en el espacio latente. La idea central consistió en proyectar los atributos de clientes y productos en un mismo espacio vectorial de baja dimensión, de manera que la similitud entre sus representaciones reflejara el grado de afinidad estimado.

Cada cliente y cada producto fueron representados a partir de un conjunto de características previamente procesadas y normalizadas. Del lado de los clientes se incluyeron variables como canal comercial, región geográfica y tamaño del punto de venta, mientras que para los productos se consideraron atributos como marca, segmento, tipo de envase y rango de precio. Estas variables fueron convertidas en vectores numéricos y luego proyectadas mediante capas densas que aprenden combinaciones no lineales entre ellas.

La arquitectura adoptada se estructuró en dos ramas simétricas: una para clientes y otra para productos. Cada rama consta de una serie de capas totalmente conectadas con activaciones ReLU y regularización mediante *dropout*, destinadas a generar embeddings de dimensión fija para cada entidad. Las salidas de ambas ramas se combinan a través de una función de similitud coseno que produce un puntaje continuo de afinidad. Este puntaje se entrena para distinguir los pares cliente-producto observados de aquellos no observados, utilizando una función de pérdida de tipo *Bayesian Personalized Ranking* (BPR) y muestreo negativo.

El modelo fue entrenado sobre el mismo conjunto de interacciones utilizado por los modelos anteriores, conservando la ventana temporal de seis meses. Se empleó un procedimiento de *mini-batch gradient descent* con optimizador Adam y tasa de aprendizaje adaptativa. Los hiperparámetros de dimensión de los embeddings, cantidad de capas y tasa de regularización se ajustaron empíricamente mediante *grid search*, evaluando el desempeño sobre una partición temporal separada. La métrica de referencia utilizada para la selección final fue el área bajo la curva ROC-AUC.

Una vez entrenado, el modelo generó representaciones vectoriales que condensan información tanto transaccional como contextual. Estas representaciones se combinaron posteriormente con los vectores aprendidos por el modelo ALS, construyendo un esquema de ensamble en el que el puntaje final se obtiene como una combinación lineal de ambos modelos. Este enfoque permitió aprovechar la capacidad del ALS para capturar patrones colaborativos y, al mismo tiempo, incorporar la flexibilidad del modelo neuronal para representar interacciones más complejas entre los atributos de clientes y productos. El resultado fue un sistema más expresivo y con mejor capacidad de generalización en contextos heterogéneos.

3.5.4. Neural Collaborative Filtering (NCF)

El modelo de *Neural Collaborative Filtering* (NCF) [29] se diseñó como una extensión híbrida del filtrado colaborativo clásico, combinando representaciones latentes aprendidas con información contextual explícita de clientes y productos. A diferencia de los enfoques lineales tradicionales, el NCF introduce una red neuronal que aprende de manera no lineal la función de interacción entre ambas entidades, permitiendo capturar relaciones más complejas y patrones de afinidad que no pueden representarse mediante una factorización matricial simple.

El modelo parte de dos vectores de entrada: el identificador de cliente y el identificador de producto, que se transforman en embeddings densos de dimensión fija aprendidos durante el entrenamiento. A estos vectores se les concatenan las características codificadas que describen el contexto de cada entidad, como canal comercial, región o tamaño del punto de venta en el caso de los clientes, y segmento, marca o tipo de envase en el caso de los productos. De esta manera, el modelo incorpora simultáneamente información colaborativa y de contenido, integrando señales de comportamiento histórico y atributos estructurales en una única representación combinada.

La arquitectura, visible en 3.10 está compuesta por una red neuronal multicapa (MLP) que recibe como entrada el vector concatenado de cliente y producto. Las capas ocultas aplican transformaciones no lineales con activaciones del tipo ReLU, reduciendo progresivamente la dimensionalidad hasta obtener un valor escalar que representa la afinidad estimada entre ambas entidades. El modelo se entrenó utilizando una función de pérdida binaria basada en la entropía cruzada, donde los pares cliente–producto con evidencia de interacción se consideraron observaciones positivas, y los pares sin historial se muestrearon como negativos. Este esquema de muestreo negativo permite que el modelo aprenda a discriminar entre combinaciones relevantes y no relevantes, ajustando la probabilidad estimada de interés para cada par.

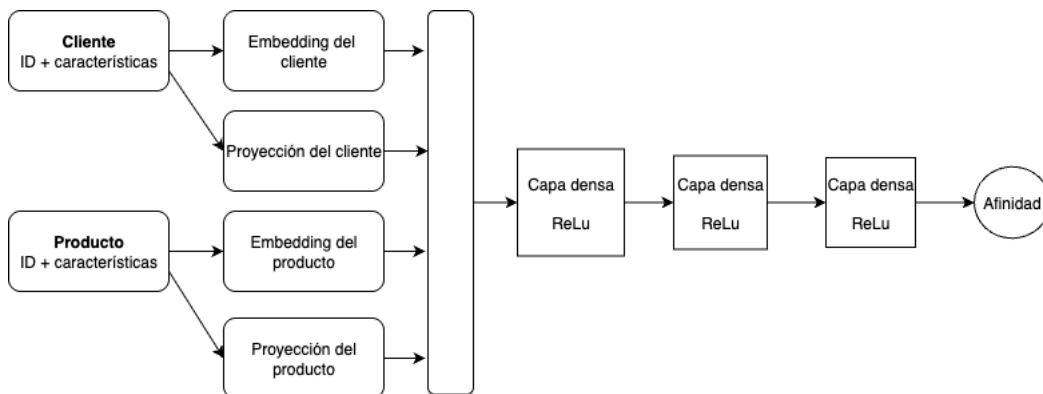


FIGURA 3.10. Arquitectura del modelo *Neural Collaborative Filtering* híbrido.

Durante el entrenamiento, se exploraron diferentes configuraciones de hiperparámetros, incluyendo la dimensión de los embeddings, la cantidad y tamaño de las capas ocultas, la tasa de aprendizaje y la proporción de muestreo negativo. El objetivo fue maximizar el área bajo la curva ROC-AUC sobre un conjunto de validación temporal, buscando un equilibrio entre capacidad predictiva y eficiencia computacional.

El modelo resultante produce un puntaje continuo de afinidad \hat{y}_{ij} para cada combinación cliente–producto. A partir de estos valores se generan listas *Top-K* personalizadas, ordenando los productos de mayor a menor probabilidad de interés. Este enfoque híbrido combina la expresividad de las redes neuronales con la capacidad de generalización de los sistemas de recomendación basados en contenido, ofreciendo una mayor flexibilidad para capturar interacciones complejas y atenuar los efectos del arranque en frío.

3.6. Implementación

La implementación del sistema de recomendación requirió articular los distintos componentes desarrollados dentro de un flujo de trabajo unificado, reproducible y escalable. Para ello se diseñó un *pipeline* modular que integra los procesos de ingestión, transformación, modelado y evaluación, con soporte para el versionado y monitoreo de artefactos en producción.

3.6.1. Diseño del pipeline de procesamiento

El flujo completo se estructuró en cuatro etapas principales: ingestión, preparación, modelado y predicción. En la fase de ingestión se integraron las fuentes de datos transaccionales, digitales y contextuales en un entorno distribuido, garantizando la consistencia de los identificadores y la alineación temporal entre registros.

Durante la preparación, se aplicaron las transformaciones de limpieza, agregación, codificación y normalización para construir la matriz cliente–producto que sirve de insumo a los modelos.

En la etapa de modelado, se ejecutaron los distintos enfoques desarrollados, almacenando sus métricas, parámetros y versiones.

Finalmente, en la fase de predicción se generaron los puntajes de afinidad y las listas *Top-K* para cada cliente, que constituyen la salida principal del sistema.

3.6.2. Integración con la infraestructura tecnológica

La ejecución del pipeline se realizó en la plataforma `Databricks` [12], que permitió procesar grandes volúmenes de datos de forma distribuida utilizando `PySpark`. Este entorno facilitó la orquestación de tareas, la paralelización de los cálculos y la trazabilidad de los resultados.

Para la gestión del ciclo de vida de los modelos se empleó `MLflow` [14], herramienta que permitió registrar los experimentos, almacenar los parámetros y métricas, y versionar los artefactos generados durante el entrenamiento. Cada ejecución de modelo quedó asociada a un identificador único, lo que posibilita reproducir resultados, comparar configuraciones y recuperar versiones históricas de los modelos entrenados.

Esta integración entre `Databricks` y `MLflow` conformó una infraestructura robusta y escalable, adecuada tanto para la experimentación iterativa como para la implementación de pipelines automatizados.

3.6.3. Estrategias de versionado y monitoreo

Con el fin de garantizar la trazabilidad del sistema, se adoptaron prácticas de control de versiones y monitoreo continuo.

El código fuente y los scripts asociados al pipeline se gestionaron mediante `GitHub` [19], lo que permitió organizar el desarrollo de manera colaborativa y mantener un historial de cambios documentado.

Por otro lado, los modelos registrados en `MLflow` se acompañaron de sus métricas de validación y fecha de generación, posibilitando un seguimiento temporal de su desempeño.

Además, se establecieron controles de consistencia sobre los datos de entrada y validaciones automáticas del formato de salida, asegurando la estabilidad operativa del sistema en cada ejecución.

En conjunto, esta arquitectura permitió implementar un flujo de trabajo integrado, auditable y escalable, garantizando la reproducibilidad de los resultados y sentando las bases para la futura incorporación de componentes en producción.

Capítulo 4

Ensayos y resultados

Todos los capítulos deben comenzar con un breve párrafo introductorio que indique cuál es el contenido que se encontrará al leerlo. La redacción sobre el contenido de la memoria debe hacerse en presente y todo lo referido al proyecto en pasado, siempre de modo impersonal.

Capítulo 5

Conclusiones

Todos los capítulos deben comenzar con un breve párrafo introductorio que indique cuál es el contenido que se encontrará al leerlo. La redacción sobre el contenido de la memoria debe hacerse en presente y todo lo referido al proyecto en pasado, siempre de modo impersonal.

Bibliografía

- [1] James Bennett y Stan Lanning. «The Netflix Prize». En: (2007). Available at: https://www.netflixprize.com/assets/GrandPrize2009_BPC_BellKor.pdf.
- [2] Jonathan L. Herlocker et al. «An Algorithmic Framework for Performing Collaborative Filtering». En: (2000), págs. 230-237. DOI: [10.1145 / 312624.312682](https://doi.org/10.1145/312624.312682).
- [3] Xinrui Zhang y Hengshan Wang. «Study on Recommender Systems for Business-To-Business Electronic Commerce». En: *Communications of the II-M A* 5.4 (2005), Article 8. DOI: [10.58729 / 1941 - 6687.1282](https://doi.org/10.58729/1941-6687.1282). URL: <https://scholarworks.lib.csusb.edu/ciima/vol5/iss4/8>.
- [4] Confidential. *[GENERAL RANK] Purchase Preference (EN)*. Inf. téc. Internal technical report, not publicly available. Confidential organization, 2023.
- [5] Francesco Ricci, Lior Rokach y Bracha Shapira. *Recommender Systems Handbook*. 2nd. Springer, 2015. DOI: [10.1007/978-1-4899-7637-6](https://doi.org/10.1007/978-1-4899-7637-6).
- [6] Gediminas Adomavicius y Alexander Tuzhilin. «Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions». En: *IEEE Transactions on Knowledge and Data Engineering* 17.6 (2005), págs. 734-749. DOI: [10.1109 / TKDE.2005.99](https://doi.org/10.1109/TKDE.2005.99).
- [7] Yifan Hu, Yehuda Koren y Chris Volinsky. «Collaborative Filtering for Implicit Feedback Datasets». En: *Proceedings of the 8th IEEE International Conference on Data Mining (ICDM)*. 2008, págs. 263-272. DOI: [10.1109 / ICDM.2008.22](https://doi.org/10.1109/ICDM.2008.22).
- [8] Badrul Sarwar et al. «Item-based Collaborative Filtering Recommendation Algorithms». En: *Proceedings of the 10th International Conference on World Wide Web (WWW)*. 2001, págs. 285-295. DOI: [10.1145/371920.372071](https://doi.org/10.1145/371920.372071).
- [9] Yehuda Koren, Robert Bell y Chris Volinsky. «Matrix Factorization Techniques for Recommender Systems». En: *Computer* 42.8 (2009), págs. 30-37. DOI: [10.1109 / MC.2009.263](https://doi.org/10.1109/MC.2009.263).
- [10] Michael J. Pazzani y Daniel Billsus. «Content-based Recommendation Systems». En: *The Adaptive Web*. Vol. 4321. Lecture Notes in Computer Science. Springer, 2007, págs. 325-341. DOI: [10.1007/978-3-540-72079-9_10](https://doi.org/10.1007/978-3-540-72079-9_10).
- [11] Paul Covington, Jay Adams y Emre Sargin. «Deep Neural Networks for YouTube Recommendations». En: *Proceedings of the 10th ACM Conference on Recommender Systems (RecSys)*. 2016, págs. 191-198. DOI: [10.1145 / 2959100.2959190](https://doi.org/10.1145/2959100.2959190).
- [12] Databricks Inc. *Databricks: Unified Data Analytics Platform*. <https://docs.databricks.com>. Official product documentation, accessed: 2025-09-28. 2024.
- [13] Matei Zaharia et al. «Resilient Distributed Datasets: A Fault-Tolerant Abstraction for In-Memory Cluster Computing». En: *Proceedings of the 9th USENIX Symposium on Networked Systems Design and Implementation (NSDI)*. 2012, págs. 2-2.
- [14] Matei Zaharia et al. «Accelerating the Machine Learning Lifecycle with MLflow». En: *IEEE Data Engineering Bulletin* 41.4 (2018), págs. 39-45.

- [15] Yifan Hu, Yehuda Koren y Chris Volinsky. «Collaborative Filtering for Implicit Feedback Datasets». En: *Proceedings of the 8th IEEE International Conference on Data Mining (ICDM)* (2008), págs. 263-272. DOI: [10.1109/ICDM.2008.22](https://doi.org/10.1109/ICDM.2008.22).
- [16] Maciej Kula. «Metadata Embeddings for User and Item Cold-start Recommendations». En: *Proceedings of the 9th ACM Conference on Recommender Systems (RecSys)*. 2015, págs. 279-282. DOI: [10.1145/2792838.2799663](https://doi.org/10.1145/2792838.2799663).
- [17] John D. Hunter. «Matplotlib: A 2D Graphics Environment». En: *Computing in Science & Engineering* 9.3 (2007), págs. 90-95. DOI: [10.1109/MCSE.2007.55](https://doi.org/10.1109/MCSE.2007.55).
- [18] Michael L. Waskom. «Seaborn: statistical data visualization». En: *Journal of Open Source Software* 6.60 (2021), pág. 3021. DOI: [10.21105/joss.03021](https://doi.org/10.21105/joss.03021).
- [19] GitHub Inc. *GitHub: Software Development Platform*. <https://github.com>. Accessed: 2025-09-28. 2024.
- [20] Richard Koch. *The 80/20 Principle: The Secret to Achieving More with Less*. Doubleday, 1998.
- [21] Chris Anderson. *The Long Tail: Why the Future of Business Is Selling Less of More*. Hyperion, 2006.
- [22] Philip Kotler y Kevin Lane Keller. *Marketing Management*. 15.^a ed. Harlow, England: Pearson Education, 2017. ISBN: 9781292092621.
- [23] Òscar Celma. *Music Recommendation and Discovery: The Long Tail, Long Fail, and Long Play in the Digital Music Space*. Springer Theses. Springer, 2010. DOI: [10.1007/978-3-642-13287-2](https://doi.org/10.1007/978-3-642-13287-2).
- [24] Liang Xiang et al. *Temporal Recommendation on Graphs via Long- and Short-Term Preference Fusion*. 2010, págs. 723-732. DOI: [10.1145/1835804.1835896](https://doi.org/10.1145/1835804.1835896).
- [25] Yehuda Koren. «Collaborative Filtering with Temporal Dynamics». En: *Communications of the ACM* 53.4 (2010), págs. 89-97. DOI: [10.1145/1721654.1721677](https://doi.org/10.1145/1721654.1721677).
- [26] Charu C. Aggarwal. *Recommender Systems: The Textbook*. Springer, 2016. DOI: [10.1007/978-3-319-29659-3](https://doi.org/10.1007/978-3-319-29659-3).
- [27] Charu C. Aggarwal. *Data Mining: The Textbook*. Springer, 2015. DOI: [10.1007/978-3-319-14142-8](https://doi.org/10.1007/978-3-319-14142-8).
- [28] Mihajlo Grbovic y Haibin Cheng. «Real-time Personalization using Embeddings for Search Ranking at Airbnb». En: (2018), págs. 311-320. DOI: [10.1145/3219819.3219885](https://doi.org/10.1145/3219819.3219885).
- [29] Xiangnan He et al. «Neural Collaborative Filtering». En: (2017), págs. 173-182. DOI: [10.1145/3038912.3052569](https://doi.org/10.1145/3038912.3052569).