.UBAfiuba
FACULTAD DE INGENIERÍA

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

Índice de figuras

Índice de tablas

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 *feedback* 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.

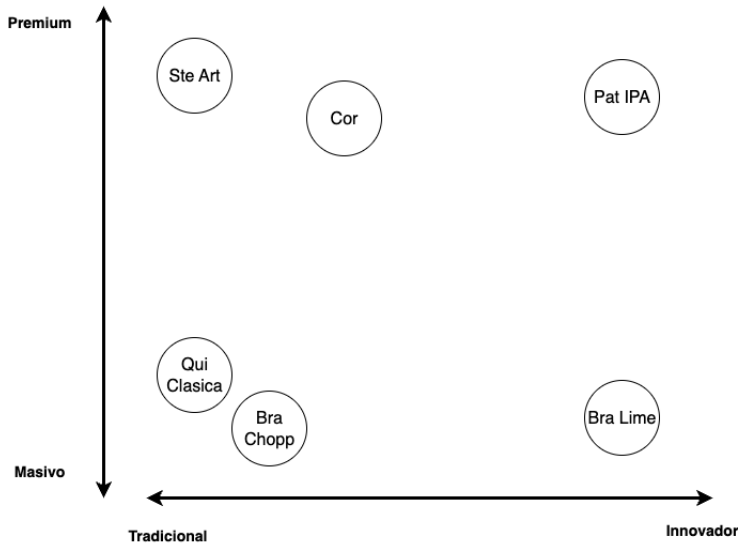

FIGURA 2.1. Ejemplo de representación de marcas de cerveza en un espacio de atributos.

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.

$$R \approx U \cdot V^T \tag{2.1}$$

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.

$$\hat{r}_{ij} = u_i \cdot v_j^T \tag{2.2}$$

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.

$$\hat{r}_{ij} = w_i \cdot x_j \tag{2.3}$$

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.

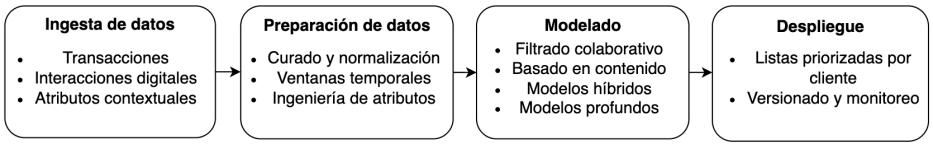

FIGURA 3.1. Arquitectura de alto nivel del sistema de recomendación.

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].

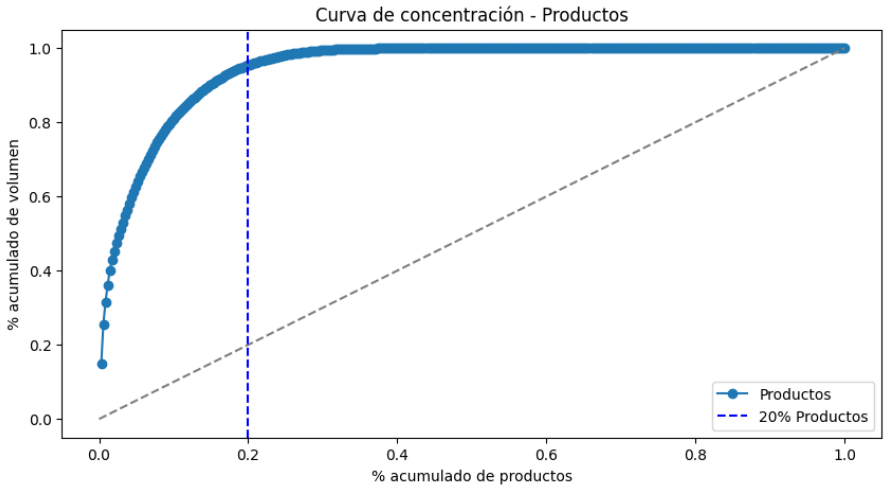

FIGURA 3.2. Concentración de productos en el portafolio.

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 generan 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.

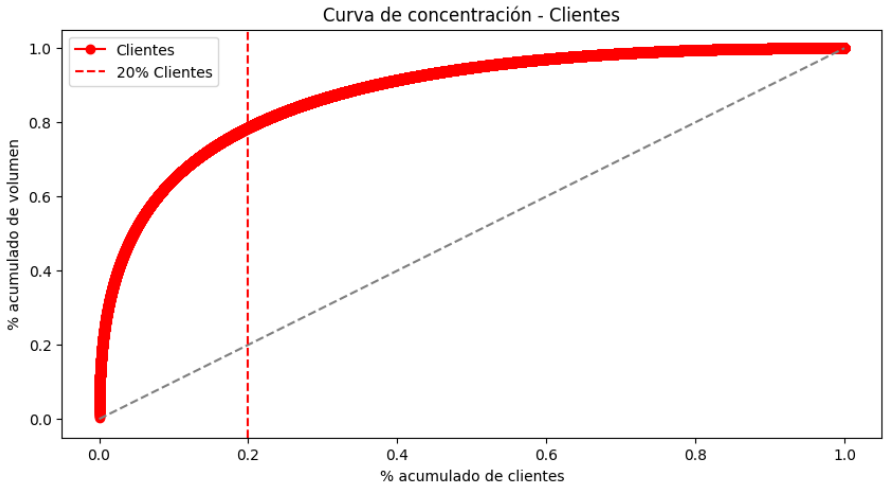

FIGURA 3.3. Concentración de clientes.

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.

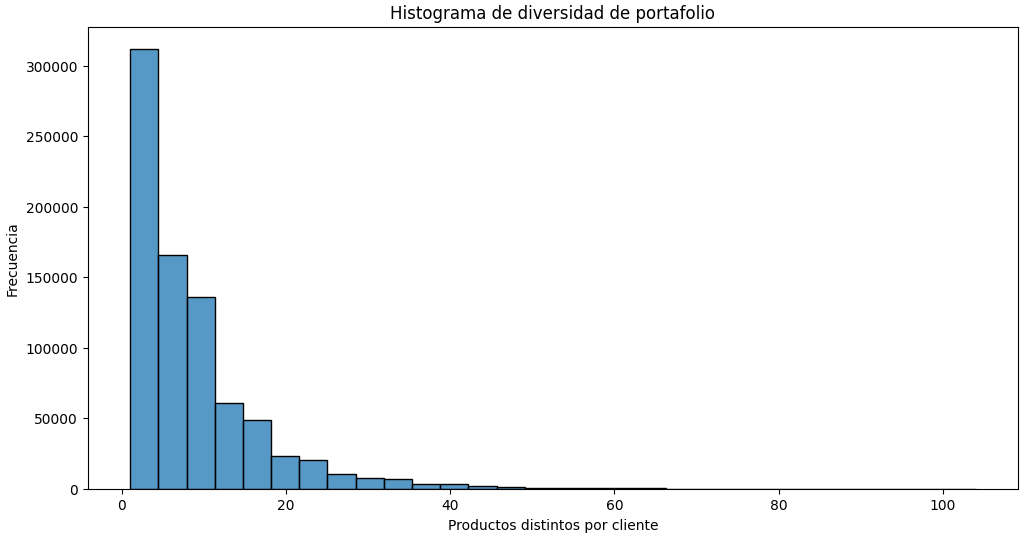

FIGURA 3.4. Histograma de diversidad de portafolio: número de productos distintos por cliente.

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. El gráfico muestra una pendiente decreciente que confirma la existencia de este

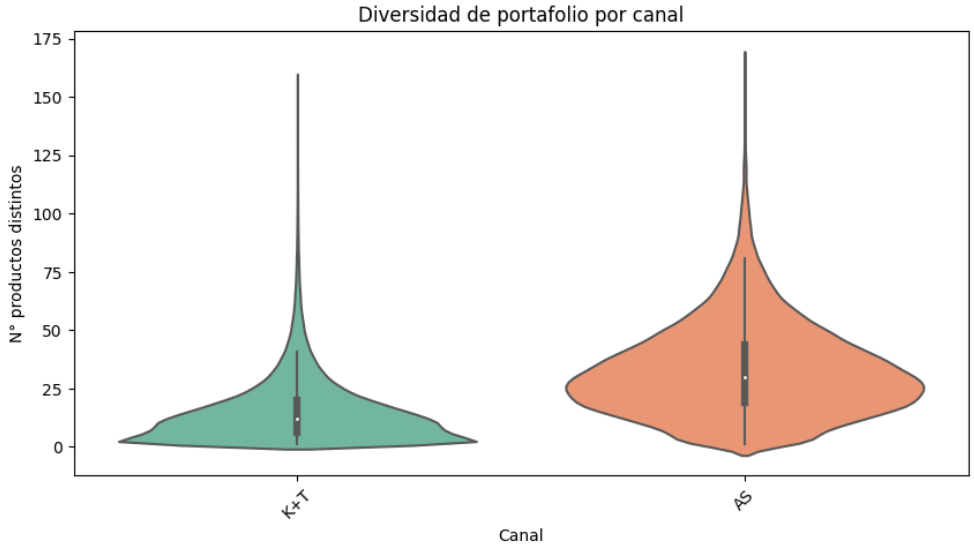

FIGURA 3.5. Diversidad de portafolio segmentada por canal comercial.

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.

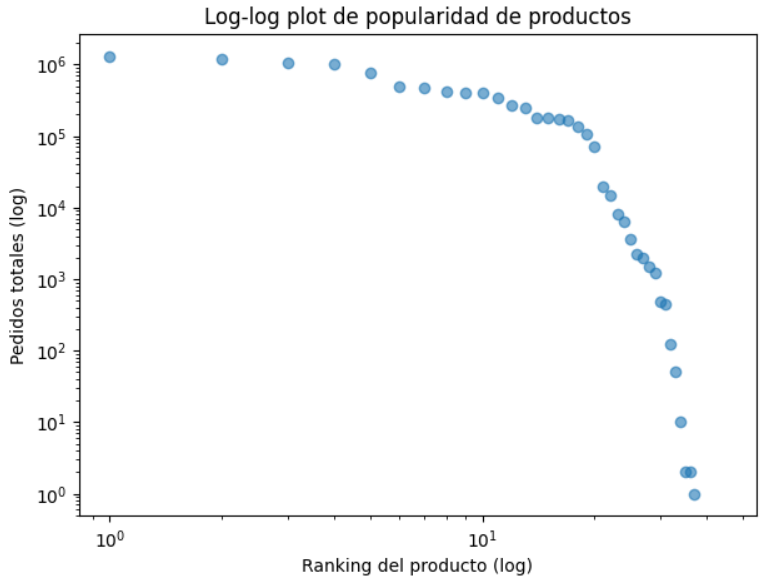

FIGURA 3.6. *Log log plot* de la popularidad de productos.

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.

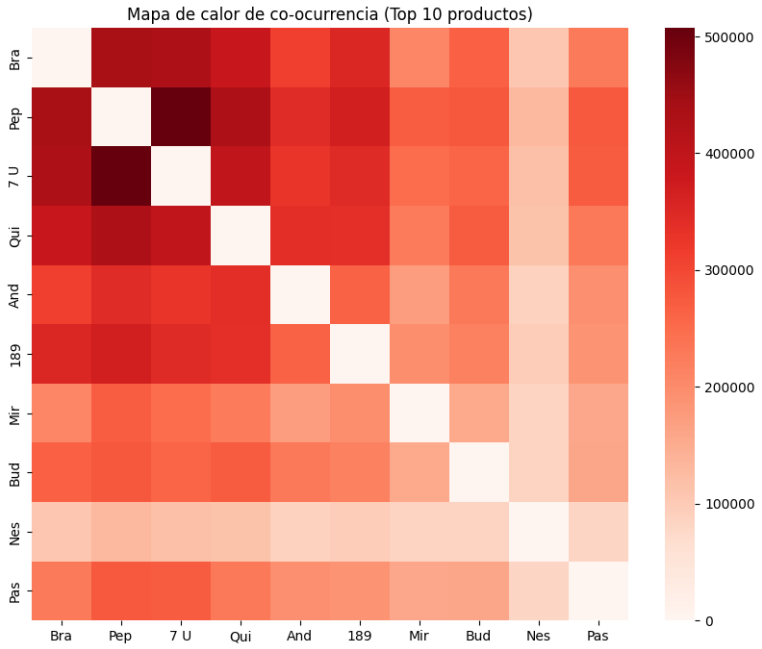

FIGURA 3.7. Mapa de calor de co-ocurrencia entre los 10 productos más relevantes.

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

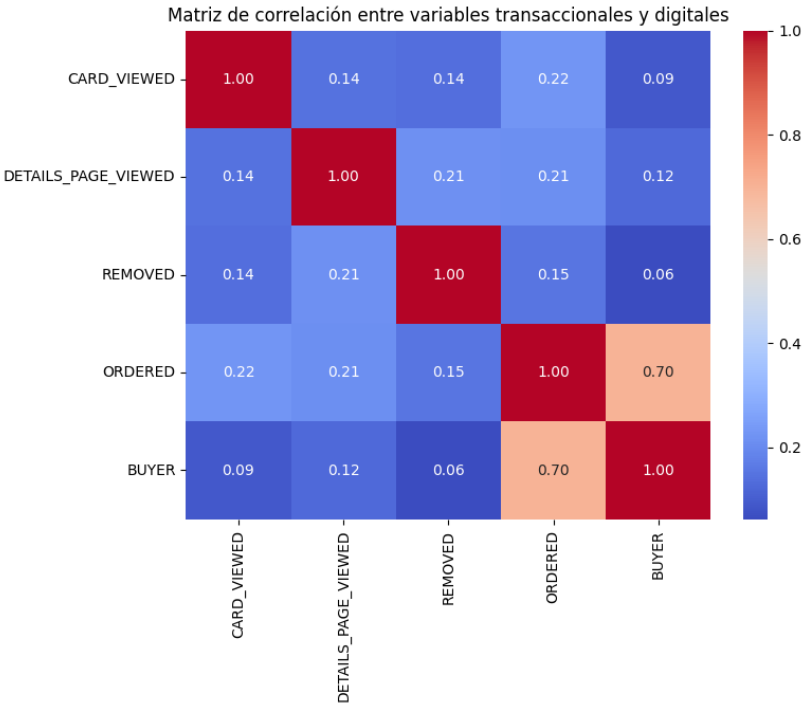

FIGURA 3.8. Matriz de correlación entre variables transaccionales y digitales.

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 (`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.

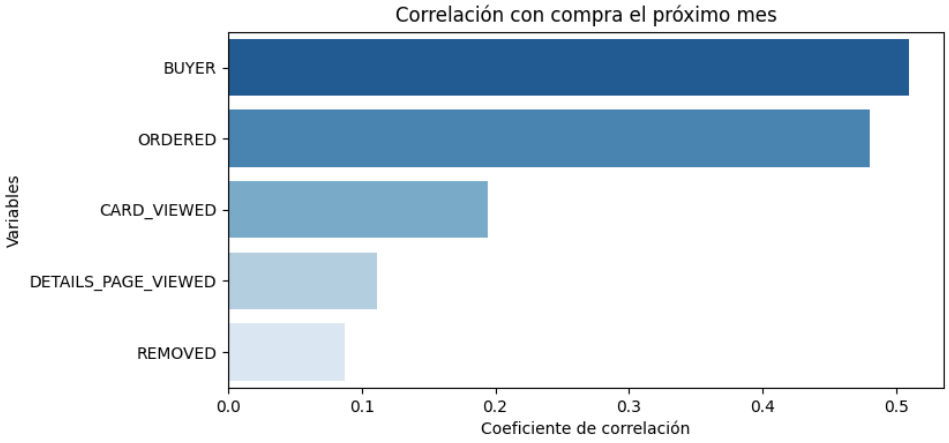

FIGURA 3.9. Correlación de variables con la compra en el mes siguiente.

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 coocurrencia 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 etapa de ingeniería de atributos constituyó uno de los pilares centrales del desarrollo del motor de afinidad, ya que definió la forma en que las distintas fuentes de datos fueron transformadas en representaciones numéricas útiles para los modelos de recomendación. En esta fase se abordó tanto la construcción de la matriz cliente–producto, que sintetiza las interacciones históricas y digitales entre ambas entidades, como la generación de atributos contextuales que describen las características estructurales de clientes y productos.

El objetivo fue capturar la complejidad del comportamiento comercial B2B en consumo masivo, donde coexisten puntos de venta con patrones de compra heterogéneos, portafolios con alta rotación y señales digitales de distinta frecuencia e intensidad. Las fuentes de información se integraron en una arquitectura analítica diseñada para producir variables consistentes, interpretables y estables en el tiempo.

3.4.1. Diseño de la matriz cliente–producto

La matriz cliente–producto constituye el núcleo del sistema de recomendación, ya que representa de forma estructurada las interacciones históricas entre los puntos de venta y el portafolio de productos. Su construcción requirió integrar información transaccional y de comportamiento digital proveniente de la aplicación BEES, transformando ambas fuentes en un conjunto de señales cuantificables que reflejan el nivel de interés o afinidad de cada cliente hacia cada producto.

Integración de fuentes y depuración de eventos

Previo a la selección de los eventos a considerar, se realizó un proceso de depuración destinado a eliminar registros residuales o no representativos. Se excluyeron interacciones que no estuvieran asociadas a un producto específico, así como aquellas que correspondían a acciones rutinarias de navegación en la aplicación y que no implicaban una manifestación de preferencia. Este filtrado fue fundamental para reducir el ruido inherente a las señales digitales, un desafío recurrente en sistemas de recomendación donde la actividad exploratoria o automatizada puede superar a la de compra efectiva [5, 24].

Una vez consolidado el conjunto de eventos válidos, la información se agrupó a nivel mensual por cliente y producto, contabilizando la cantidad de interacciones registradas para cada tipo de evento. Se seleccionaron cuatro categorías principales por su relevancia y consistencia estadística: confirmación de orden (`ORDERED`), visualización de tarjeta promocional (`CARD_VIEWED`), acceso al detalle del producto (`DETAILS_PAGE_VIEWED`) y eliminación del producto del carrito (`REMOVED`).

Los eventos `ORDERED` y `CARD_VIEWED` cuentan además con subtipos que indican el origen de la recomendación. Por ejemplo, búsquedas populares, club de descuentos, *forgotten items*, pedido fácil, compra incremental o búsquedas recientes. Estos subtipos fueron conservados, dado que reflejan distintos mecanismos de

exposición y conversión dentro de la aplicación, y demostraron poseer correlaciones diferenciales con el comportamiento de recompra y con las métricas de desempeño predictivo del modelo.

Diseño temporal y esquema de ventanas

El diseño temporal de la matriz responde a la lógica operativa del sistema: las predicciones del mes N deben generarse durante el mes $N-1$, utilizando únicamente información previa. Por ello, se implementó un esquema de ventanas móviles de seis meses, abarcando el período comprendido entre los meses $N-7$ y $N-2$. Dentro de esa ventana se calcularon tres horizontes de observación —último mes, últimos tres meses y últimos seis meses— con el fin de capturar simultáneamente señales recientes y patrones más estables.

A diferencia de los modelos supervisados tradicionales, en los sistemas de recomendación basados en *feedback* implícito no resulta necesario realizar una partición explícita entre conjuntos de entrenamiento y prueba dentro de cada ventana temporal. El objetivo no es predecir una etiqueta conocida, sino aprender representaciones latentes que describan la relación entre clientes y productos. En este contexto, el modelo se entrena utilizando la totalidad de la información disponible hasta el mes previo y se evalúa prospectivamente sobre el siguiente período, replicando las condiciones reales de operación del sistema. Este enfoque, ampliamente recomendado en contextos temporales [25, 26], evita el uso de divisiones artificiales que fragmenten el histórico, preserva la coherencia causal y permite aprovechar al máximo los datos recientes para lograr una evaluación más realista del comportamiento futuro.

Tratamiento y normalización de variables

El conjunto de variables resultante presentaba una alta heterogeneidad en magnitudes, derivada de la coexistencia de clientes con volúmenes de compra muy dispares y productos con distintos niveles de rotación. Esta disparidad podía introducir distorsiones en los cálculos posteriores, en especial en la estimación de distancias y en los procesos de normalización, ya que unos pocos valores extremos tendían a dominar la escala de las variables. Para mitigar este efecto, se aplicó una técnica de *clipping* (*winsorizing*) [27], que limita los valores por debajo del percentil 1 y por encima del percentil 99, reduciendo la influencia de observaciones atípicas sin alterar la estructura general de los datos.

Luego, cada variable fue normalizada dentro del grupo cliente–categoría, de modo que representara el peso relativo de un producto sobre el total de la categoría en el portafolio de ese cliente. Esta forma de estandarización permite que las variables reflejen proporciones comparables entre clientes de distinto tamaño o comportamiento de compra, evitando que las diferencias absolutas de escala o estacionalidad sesguen las medidas de afinidad. En consecuencia, cada *feature* captura la importancia real del producto en relación con el contexto de consumo del cliente, en lugar de su volumen bruto, lo que mejora la capacidad del modelo para generalizar patrones de preferencia.

Estimación de pesos y generación del score de preferencia

A partir de estas variables, se definieron pesos específicos para cada tipo de evento y cada ventana temporal, con el fin de obtener un score de preferencia

compuesto. La estimación de estos pesos se realizó mediante un proceso de optimización bayesiana con `Optuna` [28], maximizando la métrica *Precision@10* sobre un conjunto de validación temporal. Se ejecutaron 100 iteraciones exploratorias, observándose como patrón consistente una mayor ponderación de las señales recientes (1M >3M >6M) y una mayor relevancia de los eventos transaccionales respecto de los digitales. Los valores finales de ponderación y los resultados detallados del proceso de optimización se presentan en el apendice A.

Análisis de la matriz resultante

El resultado final fue una matriz cliente–producto donde cada celda representa un puntaje implícito de afinidad, obtenido como la combinación ponderada de todas las interacciones entre el cliente y el producto en los distintos horizontes. Esta estructura, de naturaleza dispersa, constituye la base para los algoritmos colaborativos y sirve como punto de partida para el entrenamiento de modelos como `ALS` y `LightFM`.

La matriz resultante comprendió 60,2 millones de pares cliente–producto, de los cuales el 13,5 % presentó al menos una interacción registrada, reflejando la típica estructura dispersa de los sistemas de recomendación en entornos de consumo masivo. El score de preferencia exhibió una distribución fuertemente sesgada hacia cero, con un promedio de 0,0046 y una desviación estándar de 0,0277, lo que indica que la mayoría de las combinaciones poseen niveles de afinidad bajos o nulos. Los percentiles superiores (P90 = 0,001; P99 = 0,126) confirman que las señales de alta afinidad son poco frecuentes y concentradas en un subconjunto reducido de pares.

Se observó una incidencia de arranque en frío muy acotada: 2,981 clientes (1,5 % del total) y un único producto (0,54 % de los ítems) no registraban interacciones en el período de entrenamiento. Este resultado indica una base de datos altamente densa en términos de representatividad de comportamiento, donde casi la totalidad de los clientes y productos posee algún tipo de historial previo. Aun así, la presencia de estos casos justifica la incorporación de componentes basados en contenido dentro del sistema, ya que incluso proporciones pequeñas de usuarios o ítems sin historial pueden impactar en la cobertura global de las recomendaciones [26].

Al segmentar por comportamiento de compra, los pares con transacción efectiva el proximo mes registraron un score promedio de 0,066, muy superior al de aquellos sin compra (0,0026), con una correlación de Pearson de 0,39 entre ambas variables. Este nivel de asociación, junto con un área bajo la curva ROC (AUC = 0,8766), evidencia una sólida capacidad discriminante del score para diferenciar entre pares con y sin compra [5, 29].

Un resumen de las métricas descriptivas de la matriz se presenta en la tabla 3.2, donde se evidencia su alta dispersión, baja incidencia de casos en arranque en frío y la fuerte correlación entre el score de preferencia y el comportamiento de compra observado. Estas métricas validan que la matriz sintetiza de forma adecuada la estructura latente de preferencias: preserva la dispersión y heterogeneidad propias del dominio, al tiempo que captura la intensidad relativa de las interacciones cliente–producto y su poder explicativo sobre el comportamiento de compra.

TABLA 3.2. Resumen de métricas descriptivas de la matriz cliente–producto.

Indicador	Valor
Total de pares cliente–producto	60,220,260
Densidad de la matriz (% de celdas no nulas)	13,50 %
Clientes en arranque en frío	2,981 (1,50 %)
Productos en arranque en frío	1 (0,54 %)
Media del score de preferencia	0,0046
Desvío estándar del score	0,0277
Correlación Buyer–Preference	0,3935
Área bajo la curva (AUC)	0,8766

3.4.2. Diseño de atributos de cliente y producto

Además de la matriz de interacciones, el sistema de recomendación incorpora un conjunto de atributos específicos de clientes y productos que permiten modelar afinidades a partir de sus características intrínsecas. Estos atributos complementan las señales implícitas y constituyen la base de los enfoques basados en contenido, aportando contexto y capacidad de generalización en escenarios con poca información transaccional o de arranque en frío [5, 26].

Construcción de atributos de cliente

Para los puntos de venta se construyó una representación multicomponente que sintetiza distintos aspectos del comportamiento de compra y del perfil operativo. Las variables derivadas se agruparon en torno a seis dimensiones principales: frecuencia de compra, estabilidad temporal, volumen y crecimiento, diversidad de mix, comportamiento de compra y atributos contextuales. En conjunto, estas dimensiones permiten caracterizar tanto la intensidad como la regularidad y la amplitud del consumo de cada cliente.

La frecuencia de compra se estimó a partir del número de pedidos mensuales, su media y desviación en seis meses, junto con un coeficiente de variación que distingue clientes regulares de aquellos más esporádicos. La estabilidad temporal se midió mediante la variabilidad de las ventas y el volumen transaccionado, incorporando además un indicador binario que identifica clientes estables según su dispersión relativa. El bloque de volumen incluyó métricas acumuladas de ventas y unidades totales, así como el crecimiento del último mes respecto del promedio histórico, utilizado como señal de tendencia reciente. La diversidad del mix se midió a través de la cantidad de marcas, categorías y formatos adquiridos, complementada con un índice de Shannon [30] que resume la concentración de las compras. Por su parte, las variables de comportamiento capturaron la especialización del cliente, por ejemplo, si opera únicamente con cervezas o con bebidas no alcohólicas, la proporción de productos premium en su portafolio, la continuidad de compras en meses consecutivos y la concentración del gasto en las tres principales marcas. Finalmente, las variables contextuales refieren a la geografia y canal, permitiendo diferenciar el comportamiento entre regiones y tipos de punto de venta.

Todas las variables numéricas fueron sometidas a un proceso de *clipping* (*winsorizing*) [27] para reducir la influencia de valores atípicos, seguido de una normalización por *z-score* dentro de los grupos definidos por subregión y canal. Este tratamiento permitió mantener la comparabilidad entre clientes de distinta escala y contexto operativo, evitando que los valores extremos afectaran la calibración de los modelos. Posteriormente, las variables continuas fueron discretizadas en tres niveles mediante cuantiles: bajo, medio y alto. Esto facilita su integración en modelos que utilizan representaciones categóricas o embeddings. Las variables categóricas se estandarizaron y completaron con etiquetas neutras, garantizando la consistencia del conjunto final. El resultado fue un vector descriptivo equilibrado y robusto que combina información transaccional, contextual y comportamental de cada cliente.

Construcción de atributos de producto

El diseño de atributos de producto siguió una lógica análoga. En este caso, las variables reflejan la posición del producto dentro del portafolio, su desempeño comercial y su nivel de adopción en el mercado. Se incorporaron indicadores de volumen y ventas acumuladas, número de pedidos y clientes alcanzados, crecimiento mensual y ticket promedio. La penetración se estimó como la proporción de clientes que adquirieron el producto en el período, mientras que la recurrencia representó la frecuencia media de recompra entre los compradores. También se midieron la diversificación geográfica y de canal y la composición por unidad de negocio (`CZA` o `NABS`), junto con atributos estructurales como segmento de marca, tipo de envase y condición de producto premium. Finalmente, se calcularon métricas relativas de desempeño, como la participación del producto dentro de su categoría o su marca, y un índice de diversidad de clientes que resume la dispersión del gasto a través de distintos puntos de venta.

El procesamiento estadístico aplicado a los productos replicó la metodología utilizada para los clientes, combinando *winsorizing* para el tratamiento de valores extremos, normalización por *z-score* dentro de cada unidad de negocio y discretización por cuantiles. De esta forma, se obtuvo una representación homogénea que preserva la comparabilidad entre productos de distinta rotación o escala de ventas.

Representatividad de los atributos

Con el propósito de evaluar la representatividad y calidad de la información contenida en los atributos de cliente y producto, se desarrolló un análisis exploratorio basado en métricas estadísticas y geométricas, en lugar de un modelo predictivo supervisado. Este procedimiento permitió caracterizar la varianza, la correlación con la ocurrencia de compra y la similitud estructural entre entidades, cuantificando así la diversidad informativa y la capacidad discriminante de cada conjunto de variables.

En el caso de los clientes, se analizaron 95 atributos válidos (59 numéricos y 36 categóricos), tras la conversión automática de variables textuales y la indexación de categorías. Las variables con mayor varianza correspondieron a medidas de volumen y frecuencia, como la facturación total, el número de pedidos y el volumen promedio, lo que refleja la amplia heterogeneidad de los puntos de venta.

Las correlaciones más elevadas con la variable de compra se observaron en indicadores de continuidad y madurez, tales como la cantidad de meses consecutivos con actividad y la frecuencia de pedidos, con coeficientes próximos a 0,32. La similitud coseno promedio entre representaciones de clientes fue de 0,9985, indicando una alta consistencia interna en el espacio de atributos y una concentración de la variabilidad en unas pocas dimensiones relevantes.

En los productos se evaluaron 61 atributos (37 numéricos y 24 categóricos), entre los cuales predominaron las métricas de desempeño y alcance comercial. Las variables con mayor varianza fueron las asociadas al volumen total, la facturación y el número de puntos de venta atendidos, mientras que las correlaciones más fuertes con la variable objetivo correspondieron a la penetración de clientes y a la cobertura comercial, con coeficientes superiores a 0,84. La similitud coseno promedio fue de 0,6449, evidenciando una mayor dispersión relativa respecto de los clientes, coherente con la diversidad estructural del portafolio.

Para complementar el análisis, se exploró la capacidad de las features de cliente y producto para reflejar diferencias inherentes en los datos mediante una reducción de dimensionalidad con Análisis de Componentes Principales (PCA). El objetivo fue observar si las categorías de negocio se separan naturalmente en el espacio de representación.

En el caso de los clientes, figura 3.10, la proyección PCA de los puntos de venta muestra una separación visible entre los canales Autoservicio (AS) y Kiosco + Tradicional (K+T). Si bien existe cierto solapamiento, los autoservicios tienden a concentrarse en la parte superior del plano, mientras que los puntos de venta tradicionales se agrupan hacia zonas inferiores. Esto sugiere que las features de cliente capturan patrones de comportamiento diferenciados entre canales.

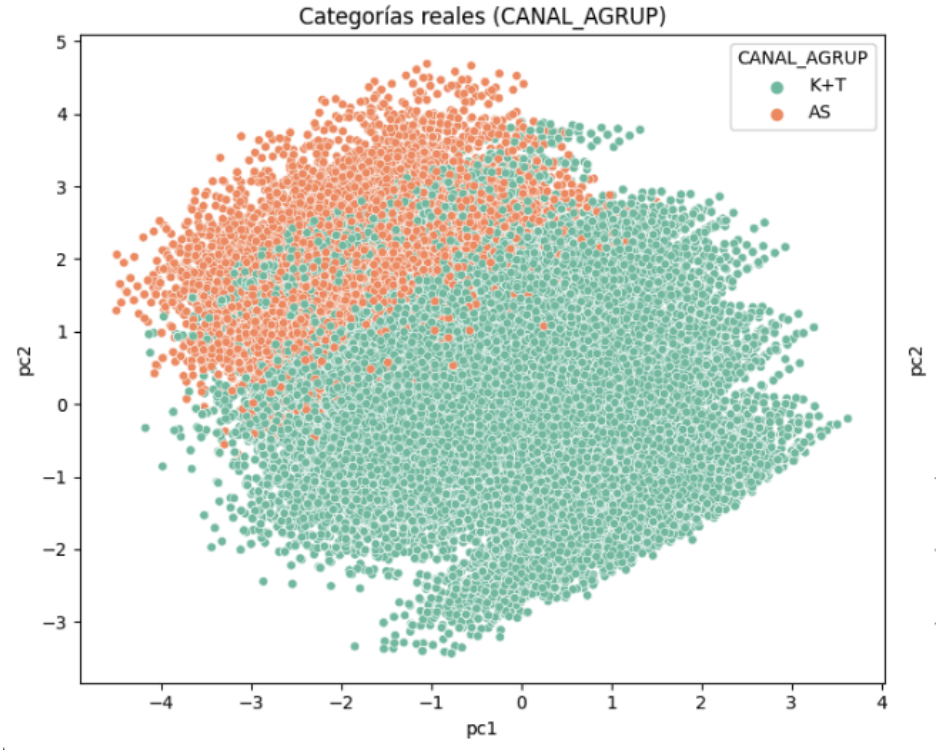

FIGURA 3.10. Proyección PCA de clientes por canal comercial.

En los productos, figura 3.11, la representación revela grupos bien definidos que corresponden a las divisiones Cerveza (CZA), No Alcohólicas (NABS) y Vinos o Adyacencias (Match). Las categorías se ubican en regiones claramente separadas del plano, lo que indica que las variables de producto describen de manera efectiva diferencias estructurales entre líneas de negocio.

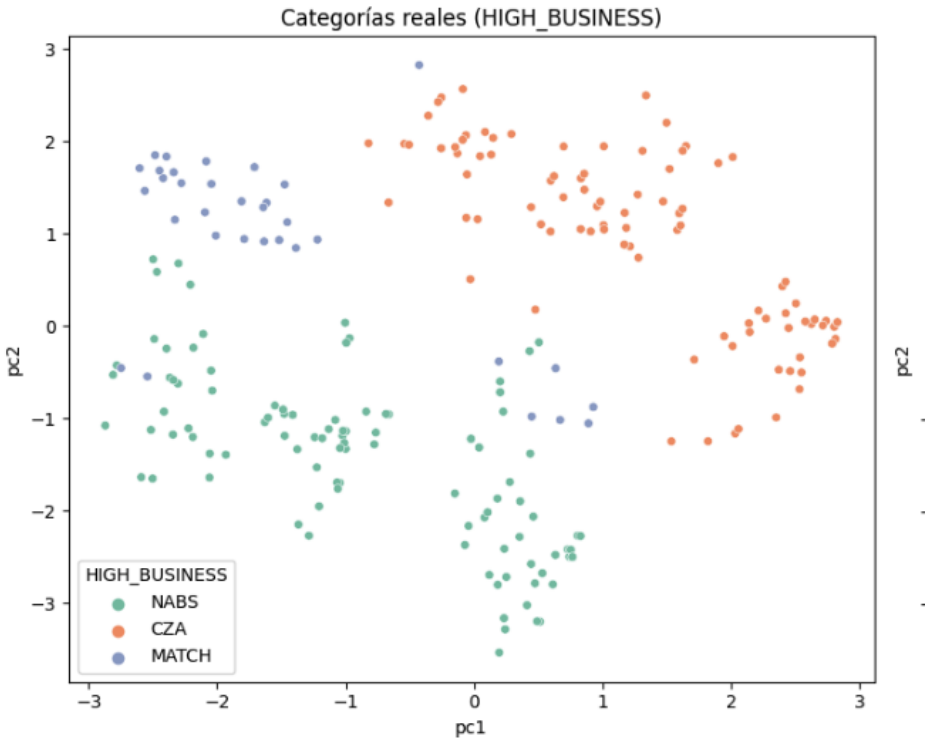

FIGURA 3.11. Proyección PCA de productos por línea de negocio.

En conjunto, los resultados del PCA demuestran que las features de producto reflejan con claridad la organización natural del portafolio, mientras que las de cliente presentan una diferenciación más moderada pero coherente con los tipos de canal. Ambas representaciones resultan adecuadas para describir la heterogeneidad comercial de la base analizada y aportan una base sólida para su integración en modelos híbridos de recomendación.

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* [31], que pueden combinarse con ALS en un esquema de ensamble. La segunda corresponde al enfoque de *Neural Collaborative Filtering* (NCF) [32], 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 enfoque desarrollado fue un modelo de filtrado colaborativo basado en factorización matricial mediante el algoritmo *Alternating Least Squares* (ALS) [9]. Este método permite descomponer la matriz cliente–producto en dos espacios latentes de menor dimensión, uno para los clientes y otro para los productos, de modo que la afinidad entre ambos se estime como el producto interno de sus vectores representativos. El objetivo es capturar patrones de coocurrencia en las interacciones históricas y proyectarlos hacia combinaciones no observadas, generando así recomendaciones personalizadas a partir del comportamiento colectivo.

Configuración del modelo

En este trabajo se utilizó la variante de ALS para *feedback* implícito, apropiada para contextos donde las señales de preferencia provienen de interacciones observadas en lugar de calificaciones explícitas [7]. Bajo este esquema, la ausencia de interacción no se interpreta como una valoración negativa, sino como falta de evidencia. La matriz de entrada corresponde al score de preferencia compuesto descrito en la sección anterior, el cual integra eventos transaccionales y digitales ponderados según su relevancia y temporalidad. Cada celda representa un grado de afinidad implícita entre el cliente y el producto, estimado a partir de seis meses de comportamiento histórico.

El modelo se configuró con un número de factores latentes ajustable (`rank`), un parámetro de regularización λ para controlar el sobreajuste y un parámetro de confianza α que pondera la influencia de las observaciones positivas frente a las ausentes [7]. El entrenamiento se realizó iterativamente, alternando la actualización de las matrices de clientes y productos hasta alcanzar convergencia.

Optimización de hiperparámetros

Para la selección de hiperparámetros se implementó un proceso de optimización bayesiana con `Optuna` [28], compuesto por veinte iteraciones orientadas a maximizar la métrica *Precision@5* sobre un conjunto de validación temporal. La búsqueda abarcó los principales parámetros del modelo: dimensión latente (`rank` $\in$

$[10, 50]$), número máximo de iteraciones (`maxIter` $\in [5, 20]$), nivel de regularización (`regParam` $\in [10^{-4}, 10^{-1}]$) y parámetro de confianza (`alpha` $\in [10^{-2}, 10]$).

En cada iteración, el modelo fue entrenado sobre las interacciones comprendidas entre los meses $N-7$ y $N-2$, y evaluado sobre el mes $N-1$, reproduciendo el flujo real de generación de recomendaciones. La métrica *Precision@5* se calculó como el promedio, por cliente, de la proporción de productos efectivamente comprados que aparecieron dentro de las primeras posiciones del ranking generado. Este criterio de optimización permitió priorizar la calidad práctica de las recomendaciones, privilegiando la capacidad del modelo para identificar los productos más relevantes en los primeros lugares de la lista.

Configuración óptima obtenida

El proceso de optimización permitió identificar la siguiente configuración como la de mejor desempeño general, maximizando la precisión en los primeros resultados sin comprometer la estabilidad del entrenamiento:

TABLA 3.3. Configuración óptima del modelo de filtrado colaborativo con `ALS` obtenida mediante optimización bayesiana.

Parámetro	**Valor óptimo**
Dimensión latente (`rank`)	26
Regularización (`regParam`)	0,0979
Iteraciones máximas (`maxIter`)	11
Restricción de no negatividad (`nonnegative`)	False
Bloques de usuario (`numUserBlocks`)	10
Bloques de ítem (`numItemBlocks`)	10
Feedback implícito (`implicitPrefs`)	True
Tamaño de bloque (`blockSize`)	4096
Parámetro de confianza (`alpha`)	0,5415
Intervalo de checkpoint (`checkpointInterval`)	10
Estrategia de arranque en frío (`coldStartStrategy`)	`drop`

Resultados y conclusiones

Los resultados obtenidos mostraron un comportamiento consistente y robusto. El modelo alcanzó una *Precision@10* de 31,5 % y *Recall@10* de 33,7 %, indicando una buena capacidad para priorizar los productos efectivamente comprados en el siguiente período. Se observó además que el ALS tiende a capturar de manera eficiente las relaciones entre clientes con historial suficiente, pero su desempeño disminuye en escenarios de arranque en frío o cuando la matriz presenta elevada dispersión. Por ello, este modelo se adoptó como línea base sobre la cual se construyeron enfoques híbridos más expresivos en las siguientes etapas.

Ell modelo ALS demostró ser una herramienta eficaz para extraer representaciones latentes de afinidad a partir del comportamiento histórico [9]. Su estructura matemática simple, estabilidad en entornos de gran escala y adecuación al feedback implícito lo convierten en un componente esencial del pipeline de recomendación desarrollado.

3.5.2. Modelo híbrido con LightFM

El segundo enfoque explorado fue un modelo *híbrido* basado en la librería `LightFM` [16], que combina técnicas de filtrado colaborativo y modelos basados en contenido dentro de un mismo marco de aprendizaje. Este modelo extiende la factorización matricial tradicional al incorporar vectores de características (*feature embeddings*) asociados tanto a los usuarios como a los ítems, permitiendo capturar información contextual incluso para aquellos pares que no poseen interacciones históricas, mitigando así el problema de arranque en frío identificado en la sección anterior.

A diferencia del ALS, que aprende representaciones exclusivamente a partir de la matriz de interacciones, LightFM incorpora atributos estructurales de clientes y productos como señales adicionales [16]. En este trabajo, las representaciones de cliente incluyeron variables de frecuencia, estabilidad, volumen, mezcla, comportamiento y contexto, mientras que las de producto comprendieron características de volumen, penetración, composición de canal, segmento, diversidad y desempeño. Cada conjunto de variables fue normalizado, discretizado y codificado antes de su integración al modelo, conforme al pipeline de ingeniería descrito previamente.

El modelo `LightFM` combina señales colaborativas y de contenido dentro de un espacio latente común, representando tanto a los clientes como a los productos mediante vectores que integran información transaccional y contextual. Su entrenamiento parte de una matriz binaria de interacciones, donde cada par cliente–producto indica la existencia o no de contacto en el período de análisis. Estas interacciones positivas constituyen la evidencia de afinidad sobre la cual el modelo aprende a distinguir entre ítems relevantes y no relevantes para cada usuario.

A diferencia del enfoque de `ALS`, que busca ajustar magnitudes continuas de preferencia implícita, `LightFM` se entrena optimizando una función de pérdida orientada al ordenamiento relativo de los ítems en el ranking. Este tipo de funciones, ampliamente utilizadas en escenarios de *feedback* implícito, penalizan los errores en las primeras posiciones y favorecen que los productos con mayor probabilidad de interacción ocupen los primeros lugares en las recomendaciones.

A cada interacción se le asignó además un peso relativo o *sample weight*, que determina su influencia durante el entrenamiento. Estos pesos se derivaron del puntaje de preferencia implícito calculado en la etapa de ingeniería de atributos y se transformaron mediante una función logarítmica que reduce la dispersión entre observaciones extremas. Posteriormente, los valores fueron normalizados alrededor de su media global, de manera que las interacciones más intensas, como compras frecuentes o múltiples eventos asociados al mismo producto, tuvieran mayor impacto que aquellas esporádicas.

Este esquema permite que el modelo capture no sólo la ocurrencia de una interacción, sino también su intensidad relativa, integrando señales de distinta fuerza en el proceso de aprendizaje. A diferencia de `ALS`, que optimiza una función cuadrática ponderada buscando reconstruir las magnitudes observadas de preferencia implícita, `LightFM` utiliza un criterio de aprendizaje orientado al ordenamiento, donde los pesos asignados a cada interacción afectan directamente la probabilidad de que un ítem sea priorizado en el ranking final [33, 16]. Esto permite capturar diferencias más finas en la intensidad de las señales, integrando

tanto la frecuencia como la relevancia relativa de cada evento dentro del proceso de entrenamiento.

El entrenamiento se llevó a cabo sobre el conjunto de interacciones ponderadas, empleando las matrices de características de usuario y producto generadas en la etapa anterior. El modelo se optimizó durante veinte épocas, utilizando procesamiento paralelo en cuatro hilos, hasta alcanzar estabilidad en la función objetivo. Esta configuración resultó adecuada para equilibrar precisión y eficiencia computacional, sirviendo como base para los distintos ensayos de complejidad incremental desarrollados a continuación.

Con el objetivo de analizar el aporte incremental de las variables contextuales, se realizaron tres configuraciones experimentales con distintos niveles de complejidad en las representaciones de usuario y producto.

Test 1 – Modelo base con contexto categórico reducido

El primer experimento con el modelo `LightFM` se diseñó como una prueba inicial para evaluar el aporte del contexto estructural más básico sobre la capacidad predictiva del sistema. En esta configuración, se incorporaron únicamente atributos categóricos que describen propiedades intrínsecas de los clientes y productos, sin incluir aún las variables derivadas del comportamiento transaccional historico.

La representación vectorial de cada cliente se construyó a partir de tres campos estructurales: la subregión, el canal de venta y un indicador binario de venta de alcohol. Estas variables permiten capturar diferencias sistemáticas entre tipos de puntos de venta en función de su localización y mix comercial, dos factores que influyen fuertemente en los patrones de compra observados.

Para los productos, se incluyeron seis descriptores fundamentales: la familia de marca, el tipo de empaque, el segmento de marca, la unidad de negocio y el indicador de contenido alcohólico. Estos campos resumen la posición del producto dentro del portafolio y su rol dentro del surtido, aspectos clave para modelar afinidades en un contexto B2B.

Todas las variables se codificaron como entidades categóricas y se proyectaron en el espacio latente del modelo mediante *feature embeddings*. El entrenamiento se realizó utilizando la función de pérdida `WARP` (*Weighted Approximate-Rank Pairwise*), orientada a optimizar directamente la posición relativa de los productos relevantes dentro del ranking de recomendaciones. Este criterio, ampliamente adoptado en contextos de *feedback* implícito [33], penaliza de forma más intensa los errores en las primeras posiciones del ranking, favoreciendo la recuperación de ítems con mayor probabilidad de interacción.

Aun con esta configuración reducida, centrada exclusivamente en variables categóricas estáticas, el modelo alcanzó una *Precision@10* de 25,3 % y *Recall@10* de 27,1 %. Esto evidencia que incluso en ausencia de señales históricas, las relaciones estructurales entre clientes y productos contienen información predictiva relevante, capaz de guiar recomendaciones con un nivel competitivo de precisión. Las recomendaciones generadas mostraron coherencia semántica, agrupando puntos de venta del mismo canal con surtidos de características similares en empaque, segmento o tipo de negocio.

En conjunto, este primer ensayo permitió validar la arquitectura híbrida de `LightFM` en su forma más simple, estableciendo una línea base a partir de la cual se incorporaron progresivamente variables discretizadas y filtros de relevancia en las siguientes configuraciones.

Test 2 – Modelo con selección explicativa de atributos discretizados

El segundo experimento extendió la configuración inicial incorporando un conjunto ampliado de atributos categóricos, seleccionados a partir de un proceso sistemático de análisis estadístico y reducción de redundancia. El objetivo fue evaluar si la inclusión de variables discretizadas derivadas del comportamiento histórico y de la estructura de mercado mejoraba la capacidad predictiva del modelo.

Para la selección preliminar de variables se aplicaron pruebas de independencia *Chi-cuadrado* sobre las variables categóricas y coeficientes de correlación de *Spearman* sobre las numéricas, priorizando aquellas con mayor fuerza de asociación con la variable objetivo de compra. Posteriormente, se eliminó la multicolinealidad mediante un umbral de correlación absoluta de $|r| > 0{,}85$, preservando únicamente las variables más representativas y no redundantes. Este procedimiento permitió conformar un subconjunto interpretativo de atributos explicativos que combinan señales de estabilidad, volumen y diversidad, sin introducir ruido ni sobreajuste.

En el caso de los clientes, se agregaron descriptores de estabilidad temporal y madurez, como la consistencia de ventas y la cantidad de meses consecutivos con actividad; medidas de especialización, como los indicadores binarios de exclusividad por unidad de negocio y variables de estructura del portafolio, incluyendo la diversidad de marcas, el número de categorías y el ticket promedio. También se incorporaron proporciones asociadas a segmentos específicos, como la participación de cervezas premium y de bebidas sin alcohol dentro del mix de cada cliente.

En el caso de los productos, se sumaron atributos relacionados con su desempeño y alcance, tales como la penetración de clientes, el volumen total vendido, la diversidad de clientes y la proporción de ventas dentro de la unidad de negocio cervecera. Todas las variables numéricas fueron previamente discretizadas en tres cuantiles (*low, medium* y *high*), asegurando comparabilidad y robustez frente a escalas heterogéneas.

El modelo se entrenó bajo la misma configuración general, utilizando la función de pérdida `WARP`. La métrica *Precision@10* de 26,1 % y el *Recall@10* de 27,8 %, mejorando significativamente el desempeño del modelo base.

El análisis cualitativo de las recomendaciones reveló una mayor capacidad de segmentación: los clientes con comportamiento estable y surtido diverso recibieron sugerencias más alineadas con su perfil, mientras que los puntos de venta especializados tendieron a recibir productos consistentes con su unidad de negocio principal. Este resultado confirma el valor de las señales discretizadas y explicativas en la construcción de representaciones híbridas, fortaleciendo la capacidad del modelo para capturar patrones de preferencia más finos sin comprometer la generalización.

Test 3 – Análisis y depuración de representaciones latentes

El tercer experimento tuvo como objetivo evaluar la calidad de las representaciones aprendidas por el modelo `LightFM` y depurar el conjunto de atributos en función de su contribución efectiva al espacio de embeddings. A diferencia de los ensayos anteriores, en esta etapa se analizaron directamente las propiedades geométricas de los vectores generados durante el entrenamiento, empleando métricas de norma, cohesión y separación para cuantificar la estructura latente subyacente.

Se diseñó un pipeline específico de evaluación de embeddings, capaz de extraer los vectores de usuario y producto almacenados en los artefactos del modelo, calcular estadísticas de dispersión y estimar su coherencia semántica. Las normas promedio de los embeddings de cliente y producto fueron de 0,031 y 0,118 respectivamente, con baja dispersión intra-grupo, lo que indica una adecuada regularización y una representación estable. La relación entre similitudes intra e inter–categoría fue de 1,17, evidenciando que los productos tienden a agruparse coherentemente dentro de sus unidades de negocio, sin perder capacidad de generalización hacia otros segmentos.

Sobre esta base se aplicó un análisis de importancia de atributos, reconstruyendo los embeddings asociados a cada *feature* explícita del modelo. La norma del vector correspondiente a cada atributo fue utilizada como indicador de relevancia latente: las características con mayor norma poseen mayor influencia en la formación del espacio de representación. Este análisis permitió identificar qué variables aportaban mayor poder discriminante y cuáles eran redundantes o marginales.

En el caso de los clientes, los atributos con mayor norma promedio correspondieron a las variables asociadas al tipo de canal comercial, la venta de productos con alcohol, la proporción de bebidas sin alcohol y la participación de cervezas premium dentro del portafolio. Estos resultados reflejan la relevancia de las señales estructurales y de composición del surtido en la caracterización de la afinidad entre puntos de venta y productos.

Entre los productos, las dimensiones más relevantes fueron la diversidad de clientes, el volumen total vendido, la cantidad de puntos de venta alcanzados y la penetración promedio, lo que evidencia una fuerte relación entre el alcance comercial y la estabilidad de la demanda en la estructura latente aprendida por el modelo.

A partir de este análisis se implementó un proceso de selección automática, conservando únicamente aquellas variables cuya norma superaba el 20 % del valor máximo dentro de su grupo y hasta dos categorías por campo. El resultado fue un conjunto final de 28 atributos de cliente y 4 de producto, conformando una base más parsimoniosa y explicable. Entre los factores retenidos se destacan las dimensiones de estabilidad temporal, diversidad de surtido y volumen de compra, que capturan aspectos complementarios del comportamiento de los puntos de venta y resultan fundamentales para modelar su propensión a interactuar con distintos productos.

La evaluación del modelo sobre el conjunto de validación temporal evidenció un desempeño estable en comparación con las configuraciones previas, con una

Precision@10 de 25,8 % y un *Recall@10* de 27,5 %. Si bien las métricas se mantuvieron en niveles similares, la depuración de atributos permitió alcanzar una representación más compacta y explicable sin comprometer la capacidad predictiva del sistema.

Este refinamiento redujo la complejidad del modelo y mejoró su interpretabilidad, al identificar de forma explícita las señales con mayor contribución a la estructura latente de afinidad. En conjunto, los resultados validan que las representaciones generadas por `LightFM` capturan de manera coherente los patrones de relación entre clientes y productos, preservando la semántica de las variables originales y estableciendo una base sólida para futuras extensiones y modelos de mayor complejidad.

Resultados comparativos de los ensayos experimentales

Los tres experimentos permitieron validar y refinar progresivamente la arquitectura híbrida basada en `LightFM`. El primer modelo, centrado en atributos estructurales, demostró que la información contextual básica es suficiente para capturar afinidades significativas entre clientes y productos. El segundo ensayo confirmó el valor de incorporar variables discretizadas de comportamiento y desempeño, mejorando la capacidad de segmentación y la precisión de las recomendaciones. Finalmente, el tercer experimento consolidó la robustez del enfoque al identificar, mediante el análisis geométrico de los embeddings, un subconjunto reducido de variables con alta relevancia latente, logrando un equilibrio óptimo entre explicabilidad y rendimiento.

En términos globales, las métricas de desempeño mostraron un comportamiento estable y competitivo entre los distintos ensayos, con valores de *Precision@10* y *Recall@10* cercanos al 26 % y 28 %, respectivamente, tal como se resume en la Tabla **??**. Estos resultados confirman la capacidad del modelo para capturar relaciones no lineales y de alta dimensionalidad en entornos con señales implícitas dispersas, ofreciendo un punto de partida sólido para su ajuste fino y extensión futura.

TABLA 3.4. Resumen de métricas de desempeño para las tres configuraciones experimentales del modelo `LightFM`.

Configuración	Precision@10	Recall@10
Test 1 – Contexto categórico reducido	25,3 %	27,1 %
Test 2 – Atributos discretizados y explicativos	26,1 %	27,8 %
Test 3 – Depuración y selección de embeddings	25,8 %	27,5 %

Optimización bayesiana de hiperparámetros

Con el objetivo de maximizar la precisión del ranking y validar la robustez del modelo frente a diferentes configuraciones, se llevó a cabo un proceso de optimización bayesiana mediante la librería `Optuna`. La búsqueda se orientó a maximizar la métrica *Precision@5*, priorizando la capacidad del sistema para ubicar los productos más relevantes en las primeras posiciones de recomendación.

El espacio de búsqueda incluyó combinaciones de número de componentes latentes (`no_components` $\in \{32, 64, 96, 128, 192\}$), funciones de pérdida `WARP` y

BPR, tasas de aprendizaje en el rango $[5 \times 10^{-4}, 5 \times 10^{-3}]$, y parámetros de regularización item_alpha y user_alpha en el intervalo $[10^{-6}, 10^{-4}]$. También se evaluaron valores de max_sampled entre 5 y 15, buscando un balance adecuado entre exploración y costo computacional.

El proceso permitió identificar la configuración con mejor desempeño general, manteniendo la estabilidad del entrenamiento y optimizando la calidad del ranking en escenarios reales de recomendación. La configuración óptima encontrada se detalla en la tabla 3.5.

TABLA 3.5. Configuración óptima del modelo híbrido con LightFM obtenida mediante optimización bayesiana.

Parámetro	**Valor óptimo**
Función de pérdida (loss)	bpr
Dimensión latente (no_components)	96
Tasa de aprendizaje (learning_rate)	0,00489
Regularización de ítems (item_alpha)	$3{,}48 \times 10^{-5}$
Regularización de usuarios (user_alpha)	$2{,}23 \times 10^{-6}$
Muestras negativas máximas (max_sampled)	8
Algoritmo de optimización (learning_schedule)	adagrad

Con esta configuración, el modelo alcanzó una *Precision@10* de **29,2 %** y un *Recall@10* de **31,2 %**, superando las versiones anteriores tanto en precisión como en cobertura, y consolidándose como la mejor alternativa dentro del conjunto de modelos evaluados.

Estos resultados confirman que la combinación de la función de pérdida BPR con un número intermedio de componentes latentes ofrece un equilibrio óptimo entre capacidad representacional y generalización, maximizando la recuperación de productos relevantes sin sobreajustar a las interacciones más frecuentes. El modelo resultante constituye la versión final del motor híbrido de afinidad, integrando señales transaccionales y contextuales dentro de un espacio latente de alta coherencia semántica.

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) [32] 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 3.12 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 negativas. 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.

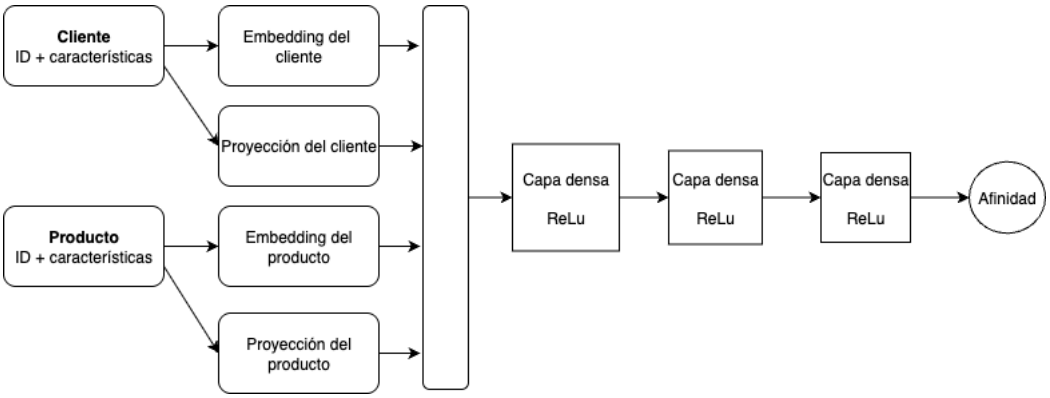

FIGURA 3.12. Arquitectura del modelo *Neural Collaborative Filtering* híbrido.

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.

Apéndice A

Optimización de pesos por evento y ventana temporal

Este anexo documenta los resultados del proceso de optimización bayesiana realizado con `Optuna`, cuyo objetivo fue estimar los pesos relativos de cada tipo de evento y de cada ventana temporal para la construcción del *score* de preferencia cliente-producto. El procedimiento se orientó a maximizar la métrica *Precision@10* sobre un conjunto de validación temporal, garantizando un balance adecuado entre la relevancia de las señales recientes y la estabilidad de los patrones históricos.

A.1. Estrategia de optimización

La optimización se ejecutó sobre 100 iteraciones mediante búsqueda bayesiana. En cada *trial* se ajustaron simultáneamente los pesos temporales y los pesos por tipo de evento (α_e), aplicándolos en la generación del *score* compuesto de afinidad. La métrica de desempeño se calculó utilizando ventanas móviles de seis meses, entrenando con el histórico comprendido entre los meses $N-7$ y $N-2$, y validando la capacidad predictiva sobre el mes $N-1$.

El resultado final mostró un patrón consistente: las señales más recientes (1M) aportan mayor información predictiva que las históricas (6M), y los eventos transaccionales tienden a tener mayor peso que los digitales, especialmente aquellos asociados a recomendaciones personalizadas o a flujos de compra recurrentes.

A.2. Pesos óptimos por ventana temporal

TABLA A.1. Pesos óptimos de las ventanas temporales.

Ventana	Peso
1 mes (reciente)	0.3925
3 meses (intermedia)	0.4735
6 meses (larga)	0.1340

Los pesos asignan mayor relevancia a los comportamientos más recientes, reflejando la mayor capacidad predictiva de las señales cercanas en el tiempo.

A.3. Pesos óptimos por tipo de evento

TABLA A.2. Pesos óptimos por tipo de evento (*event_weights*).

Evento	Peso (α_e)
BUYER	0.2208
ordered_QUICK_ORDER	0.1170
ordered	0.0700
card_viewed_QUICK_ORDER	0.1318
card_viewed_FORGOTTEN_ITEMS	0.0630
details_page_viewed	0.0066
card_viewed_CROSS_SELL_UP_SELL	0.0916
ordered_CROSS_SELL_UP_SELL	0.1882
ordered_FORGOTTEN_ITEMS	0.2174
ordered_RECENT_SEARCHES	0.0734
ordered_CLUB_B	0.1008
ordered_POPULAR_SEARCHES	0.1896
removed	0.1228
card_viewed_RECENT_SEARCHES	0.1786
card_viewed_POPULAR_SEARCHES	0.2028
card_viewed	0.0132
card_viewed_CLUB_B	0.0128

A.4. Análisis e interpretación

Los pesos reflejan una jerarquía coherente con el proceso de compra en la plataforma BEES: las órdenes efectivas (`BUYER`, `ordered_*`) constituyen las señales más predictivas de recompra, seguidas por las interacciones promocionales y de exposición de producto (`card_viewed_*`). Las categorías vinculadas a mecanismos de recomendación específicos, como *Cross-Sell/Up-Sell* y *Forgotten Items*, presentan una fuerte correlación con la conversión, lo que respalda su priorización en el cálculo del *score* final.

En contraste, los eventos de exploración (`details_page_viewed`) o de fricción (`removed`) tienen menor peso, lo que confirma que su valor informativo es limitado cuando se los considera de forma aislada.

El conjunto de ponderaciones resultante fue utilizado para construir el *score* de preferencia compuesto en la ecuación A.1:

$$\text{Score}_{ij} = \sum_{e \in E} \sum_{w \in W} \alpha_e \cdot \beta_w \cdot x_{ij}^{(e,w)} \tag{A.1}$$

donde α_e representa el peso por tipo de evento y β_w el peso temporal. Este valor resume la intensidad y actualidad de las interacciones del cliente i con el producto j, constituyendo el insumo principal de la matriz cliente-producto empleada en los modelos de recomendación.

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.

[3] Xinrui Zhang y Hengshan Wang. «Study on Recommender Systems for Business-To-Business Electronic Commerce». En: *Communications of the IIMA* 5.4 (2005), Article 8. DOI: 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.

[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.

[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.

[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.

[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.

[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.

[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.

[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.

[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.

[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.

[18] Michael L. Waskom. «Seaborn: statistical data visualization». En: *Journal of Open Source Software* 6.60 (2021), pág. 3021. DOI: 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] Oscar 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.

[24] Shuai Zhang et al. «Deep learning based recommender system: A survey and new perspectives». En: *ACM Computing Surveys* 52.1 (2019), págs. 1-38.

[25] Yehuda Koren. «Collaborative Filtering with Temporal Dynamics». En: *Communications of the ACM* 53.4 (2010), págs. 89-97. DOI: 10.1145/1721654.1721677.

[26] Charu C. Aggarwal. *Recommender Systems: The Textbook*. Springer, 2016. DOI: 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.

[28] Takuya Akiba et al. «Optuna: A next-generation hyperparameter optimization framework». En: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (2019), págs. 2623-2631.

[29] Xiangyu Zhao, Julian McAuley et al. «Learning from implicit feedback: Evaluating recommender systems with AUC and beyond». En: *Information Retrieval Journal* 25.3 (2022), págs. 345-368.

[30] Claude E. Shannon. «A Mathematical Theory of Communication». En: *Bell System Technical Journal* 27.3 (1948), págs. 379-423. DOI: 10.1002/j.1538-7305.1948.tb01338.x.

[31] 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.

[32] Xiangnan He et al. «Neural Collaborative Filtering». En: (2017), págs. 173-182. DOI: 10.1145/3038912.3052569.

[33] Steffen Rendle et al. «BPR: Bayesian Personalized Ranking from Implicit Feedback». En: (2009), págs. 452-461. URL: https://arxiv.org/abs/1205.2618.