

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, Septiembre de 2025

Resumen

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

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

Agradecimientos

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

Índice general

Resumen	I
1. Introducción general	1
1.1. Marco de la propuesta	1
1.2. Definición del problema	2
1.3. Estado del arte	4
1.3.1. Referencias en sistemas de recomendación	4
1.3.2. Sistemas de recomendación en B2B	4
1.3.3. Caso de implementación	5
1.3.4. Lecciones aprendidas	5
1.4. Motivación	6
1.5. Objetivos y alcance	7
2. Introducción específica	9
2.1. Sistemas de recomendación	9
2.1.1. Funcionamiento de los sistemas de recomendación	9
2.1.2. Tipos de <i>feedback</i>	10
2.1.3. Filtrado colaborativo	10
2.1.4. Sistemas basados en contenido	11
2.2. Fuentes de información	11
2.3. Herramientas utilizadas	12
3. Diseño e implementación	13
4. Ensayos y resultados	15
5. Conclusiones	17
Bibliografía	19

Índice de figuras

2.1. Ejemplo de representación de marcas de cerveza en un espacio de atributos.	10
---	----

Índice de tablas

1.1. Ventajas y desventajas de enfoques en recomendación	6
--	---

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, 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 *Business to Customer* (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.

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

1.3.3. Caso de implementación

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

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

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

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

1.3.4. Lecciones aprendidas

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

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

TABLA 1.1. Ventajas y desventajas de los enfoques revisados.

Enfoque / Caso	Ventajas principales	Desventajas principales
Benchmarks B2C (Netflix, Amazon, etc.)	Alta precisión y escalabilidad. Abundancia de datos y señales digitales. Estándares de evaluación consolidados.	Contextos con abundancia de <i>feedback</i> explícito/implícito, poco comparables al B2B. No consideran objetivos de negocio específicos.
Literatura B2B	Reconoce particularidades de clientes empresariales. Identifica beneficios en reducción de costos y fortalecimiento de relaciones.	Pocas implementaciones reales. Escasa estandarización de métricas y <i>datasets</i> . 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.

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, 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 en los datos disponibles. 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.

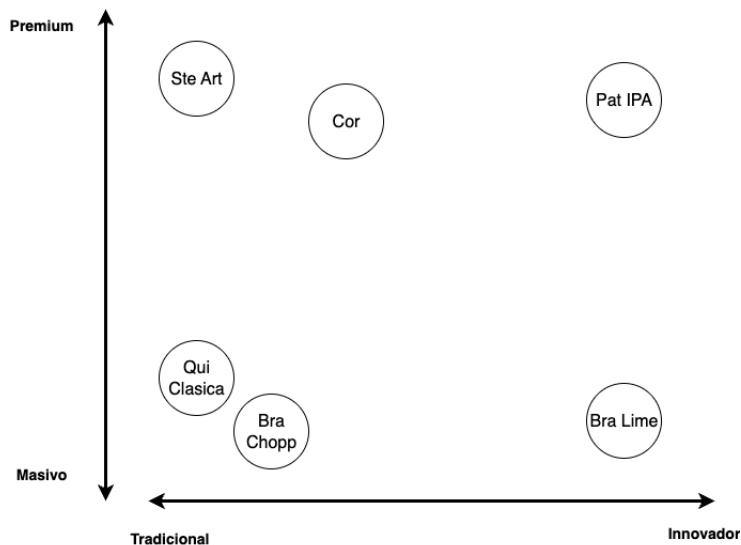


FIGURA 2.1. Ejemplo de representación de marcas de cerveza en un espacio de atributos.

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. 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

Entre los enfoques más utilizados se encuentra el filtrado colaborativo, que parte de la hipótesis de que “usuarios similares tienden a preferir ítems similares”. Este enfoque puede abordarse desde dos variantes principales. 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:

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

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

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

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

2.1.4. Sistemas basados en contenido

Otro enfoque ampliamente utilizado es el de los sistemas basados en contenido, que centran la recomendación en las características de los productos y en el perfil de cada cliente. 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:

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

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

2.2. Fuentes de información

El desarrollo del sistema de recomendación se apoya en un conjunto diverso de datos que, en su integración, permiten construir una representación enriquecida de la relación cliente–producto. Estas fuentes incluyen información transaccional, eventos de interacción en la aplicación BEES y atributos contextuales de clientes y productos. En conjunto, estos elementos aportan evidencia implícita de interés y afinidad, y constituyen la base para la estimación de scores personalizados.

Los datos transaccionales reflejan las compras efectivamente realizadas por cada cliente. Dentro de esta categoría se consideran las siguientes variables: volumen adquirido por producto, frecuencia de compra, recencia de la última operación, monetización total y participación relativa de cada producto en el mix de compras del cliente. Estas medidas permiten capturar el grado de preferencia y relevancia actual de cada artículo en la cartera de cada punto de venta.

A su vez, se incorporan eventos de comportamiento generados en la aplicación BEES, los cuales funcionan como señales implícitas de interés. Entre ellos se incluyen agregados y remociones de productos del carrito, visualizaciones de fichas de producto, búsquedas realizadas, calificaciones o reseñas, y clics en promociones. Estos registros aportan información complementaria sobre el proceso de consideración del cliente, incluso en los casos en los que no se concreta una transacción.

La caracterización se completa con variables de contexto asociadas tanto a clientes como a productos. Del lado del cliente, se integran atributos como el canal comercial y la ubicación geográfica. Del lado del producto, se consideran propiedades como marca, segmento, calibre y participación de mercado. Este conjunto de información contextual permite capturar heterogeneidades relevantes que condicionan la relación de afinidad.

De esta forma, la combinación de datos transaccionales, eventos digitales y atributos contextuales permite construir una representación integral de las interacciones cliente–producto, sobre la cual se sustenta el modelado de afinidad.

2.3. Herramientas utilizadas

El desarrollo del trabajo se apoyó en un conjunto de herramientas tecnológicas que permitieron gestionar de manera eficiente el ciclo de vida completo del sistema de recomendación, desde la preparación de datos hasta el despliegue de modelos.

En primer lugar, se utilizó la plataforma *Databricks*, que combina capacidades de procesamiento distribuido con un entorno colaborativo para el análisis de datos. *Databricks* permitió integrar múltiples fuentes, realizar transformaciones a gran escala con *PySpark* y organizar flujos de trabajo de forma reproducible y escalable. Su uso facilitó tanto la exploración inicial como la construcción de pipelines de ingeniería de atributos.

En segundo lugar, se empleó *MLflow* como herramienta de gestión del ciclo de vida de modelos. A través de esta plataforma se registraron experimentos, parámetros, métricas y versiones de modelos, lo que aseguró trazabilidad y comparabilidad entre distintos enfoques. Asimismo, se utilizaron las funcionalidades de almacenamiento y versionado de artefactos para garantizar la reproducibilidad de resultados y la posibilidad de mantener un repositorio consolidado de modelos entrenados.

El trabajo también se apoyó en librerías de aprendizaje automático ampliamente utilizadas en la comunidad científica y profesional. Entre ellas se destaca *MLlib*, utilizada para implementar la factorización matricial con *ALS*. Asimismo, se empleó *LightFM* para el desarrollo de modelos híbridos de filtrado colaborativo. Finalmente, se incorporó *PyTorch* como entorno de *deep learning*, lo que permitió construir arquitecturas más complejas capaces de capturar patrones no lineales en las interacciones cliente–producto. Adicionalmente, se utilizaron librerías de visualización como *matplotlib* y *seaborn* para generar representaciones gráficas que complementaron el análisis exploratorio y la evaluación de resultados.

Finalmente, se recurrió a entornos de control de versiones y colaboración, en particular *GitHub*, lo que permitió organizar el código en repositorios estructurados, registrar cambios de manera sistemática y facilitar la integración de componentes en distintas fases del trabajo.

En conjunto, estas herramientas proporcionaron una infraestructura sólida para desarrollar, evaluar y documentar el sistema de recomendación, lo que aseguró tanto la calidad técnica como la escalabilidad del enfoque implementado.

Capítulo 3

Diseño e implementación

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 4

Ensayos y resultados

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

Capítulo 5

Conclusiones

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

Bibliografía

- [1] James Bennett y Stan Lanning. «The Netflix Prize». En: (2007). Available at: https://www.netflixprize.com/assets/GrandPrize2009_BPC_BellKor.pdf.
- [2] Jonathan L. Herlocker et al. «An Algorithmic Framework for Performing Collaborative Filtering». En: (2000), págs. 230-237. DOI: [10.1145/312624.312682](https://doi.org/10.1145/312624.312682).
- [3] Xinrui Zhang y Hengshan Wang. «Study on Recommender Systems for Business-To-Business Electronic Commerce». En: *Communications of the IIMA* 5.4 (2005), Article 8. DOI: [10.58729/1941-6687.1282](https://doi.org/10.58729/1941-6687.1282). URL: <https://scholarworks.lib.csusb.edu/ciima/vol5/iss4/8>.
- [4] Lucas Neto Nakadaira and Matheus Boccaletti de Paula and Joao Victor Barbosa Silva. *[GENERAL RANK] Purchase Preference (EN)*. Inf. téc. Internal report, not publicly available. Ambev - BEES Digital Platform, 2023.