Tipo de documento: Tesis de maestría

Escuela de Negocios. Master in Management + Analytics

Un enfoque de aprendizaje automático para predicción de demanda de productos sin historial de ventas

Autoría: Brusco, Victoria
Año: 2024

UNIVERSIDAD TORCUATO DI TELLA

MASTER IN MANAGEMENT + ANALYTICS TESIS

UN ENFOQUE DE APRENDIZAJE AUTOMÁTICO PARA PREDICCIÓN DE DEMANDA DE PRODUCTOS SIN HISTORIAL DE VENTAS

TESIS

Victoria Brusco

Mayo 2024

Tutor: Josefina Dalla Via

Agradecimientos

Quiero expresar mi más sincero agradecimiento a todas las personas que hicieron posible esta tesis.

A mi tutora de tesis, Josefina Dalla Via, un agradecimiento muy especial por tu guía y valiosos consejos durante todo este proceso. Sin tu experiencia y dedicación, este trabajo no hubiera sido posible.

A Juan Manuel Majul, mi más profundo reconocimiento por tu colaboración y empuje desde el inicio de esta investigación. Tu conocimiento y constante compromiso fueron invaluables. Gracias por estar siempre dispuesto a ofrecer tu ayuda.

Por último, extiendo mi gratitud a todas las personas, incluyendo familia y amigos, que de una manera u otra contribuyeron a que este proyecto se hiciera realidad. A todos ustedes, muchas gracias.

Resumen

Uno de los pilares de una gestión de inventario eficiente es el equilibrio entre mantener un nivel de *stock* adecuado para satisfacer la demanda y evitar costos operativos excesivos. Gracias al análisis de datos, que ha crecido en las últimas décadas, se pueden agilizar los procesos de gestión en la industria logística. Al analizar grandes volúmenes de datos, las empresas logísticas pueden optimizar la gestión de inventarios y reducir costos de transporte[1].

A pesar de estos avances, aún persisten desafíos en la toma de decisiones basada en datos. Un ejemplo es el *marketplace* más grande de América Latina, que no solo ofrece una plataforma de comercio electrónico, sino que también proporciona una variedad de servicios logísticos a sus vendedores. Uno de estos servicios, conocido como *fulfillment*, permite que el *marketplace* almacene los productos de los vendedores en sus propios almacenes y se encargue del empaquetado y envío cuando se realiza una compra.

El núcleo del problema radica en la capacidad de pronosticar la demanda con precisión para determinar cuánto inventario debe almacenar cada usuario para cada producto, considerando la limitación de espacio disponible. Para abordar esta cuestión, la empresa ha empleado algoritmos de aprendizaje automático que analizan las ventas de los últimos noventa días de cada producto y estiman la demanda proyectada para las próximas semanas.

Ahora bien, ¿qué ocurre con los productos nuevos que no tienen historial de ventas pasadas? En estos casos, la empresa ha establecido un límite máximo de 20 unidades que cada vendedor puede almacenar de un producto en particular, tanto en México como en Brasil. Este umbral es definido arbitrariamente y no considera ningún atributo del producto ni su naturaleza intrínseca.

En el presente trabajo se argumenta que, utilizando técnicas de aprendizaje automático con el historial de productos de características similares, se puede optimizar esta decisión. Para lograrlo, se ha enfocado en la vertical de electrónica de consumo, específicamente en el segmento de celulares y computadoras, que representa el 35% de la facturación mensual del servicio de *fulfillment*. La meta consiste en determinar si el umbral de 20 unidades es el más adecuado o si, por el contrario, se subestima o sobrestima la demanda potencial.

Con este fin, se aplicaron modelos de *machine learning* para predecir la demanda de productos nuevos, considerando atributos inherentes y patrones de otros productos similares. Utilizando técnicas de aprendizaje supervisado y diversos algoritmos, los resultados han sido prometedores. Un modelo destacó

con un *F1-score* de 0.92 en datos de prueba, sugiriendo que ajustar el límite de unidades almacenadas podría reducir costos y aumentar ganancias, con una potencial disminución de hasta el 10% en el número de unidades almacenadas en los centros de distribución del *marketplace*.

Este enfoque ofrece una solución más dinámica y basada en datos que la actual regla fija, mejorando la eficiencia en la gestión de inventarios y contribuyendo significativamente a la rentabilidad del *marketplace*.

Abstract

One of the fundamental pillars to achieve efficient inventory management lies in the delicate balance between two crucial factors: the need to maintain an adequate level of stock to meet demand and the prevention of excessive operational costs associated with this storage. Thanks to the growing data analysis in recent decades, new possibilities are opening up to streamline management processes in the logistics industry. By analyzing large volumes of data, logistics companies can gain valuable insights into their operations and identify areas for improvement, from optimizing inventory management to reducing transportation costs.

Despite these advancements, challenges persist in the realm of data-driven decision making. Such is the case with the largest marketplace in Latin America, which not only offers an e-commerce platform but also provides a variety of logistic services to its sellers. One of these services, known as fulfillment, allows the marketplace to store sellers' products in its own warehouses and handle the packaging and shipping process when a purchase is made.

The core of the problem lies in the ability to accurately forecast demand to determine how much inventory each user should store for each product, considering the limitation of available space. To address this issue, the company has employed machine learning algorithms to analyze the sales of the last ninety days of each product and estimate the projected demand for the upcoming weeks.

Now, what happens with new products that have no past sales history? In such cases, the company has established a maximum limit of 20 units that each seller can store for a particular product, both in Mexico and Brazil. This threshold is arbitrarily defined and does not consider any attributes of the product or its intrinsic nature.

This work argues that by using machine learning techniques with the history of products with similar characteristics, this decision can be optimized. To achieve this, the focus has been on the consumer electronics vertical, specifically in the segment of cell phones and computers, which represents about 35% of the monthly billing of the fulfillment service. The goal is to determine whether the 20-unit threshold is the most appropriate or if, on the contrary, it underestimates or overestimates potential demand.

To this end, machine learning models were applied to predict the demand for new products, considering inherent attributes and patterns of other similar products. Using supervised learning techniques and various algorithms, the results have been promising. One model stood out with an *F1-score* of 0.92 on test

data, suggesting that adjusting the limit of stored units could reduce costs and increase profits, with a potential reduction of up to 10% of units in the marketplace's distribution centers.

This approach offers a more dynamic and data-based solution than the current fixed rule, improving efficiency in inventory management and significantly contributing to the profitability of the marketplace.

Índice

1. Introducción

La sección examina la evolución del comercio electrónico y el auge de los marketplaces en el contexto de la pandemia, resaltando el crecimiento exponencial del e-commerce y los desafíos asociados a los distintos tipos de logística ofrecidos por las plataformas. Se identifica el problema de la incertidumbre en la demanda de nuevos SKU en uno de los marketplaces más grandes de la región, que conduce a políticas de ingreso de inventario en fulfillment ineficientes. Se presenta el objetivo de la tesis, que constan en desarrollar un modelo de machine learning basado en productos similares para pronosticar la demanda de nuevos productos en electrónica de consumo, con el fin de optimizar la gestión de inventario y reemplazar la práctica actual de definir arbitrariamente el stock permitido en fulfillment.

En el contexto de 2020, caracterizado por el cierre de negocios físicos y una marcada tendencia hacia lo digital, la digitalización experimentó un crecimiento exponencial a nivel global. Las restricciones impuestas por la pandemia llevaron a que aproximadamente 150 millones de personas realizaran su primera compra en línea[2], resultando en un incremento del 26.4% en el comercio electrónico global, alcanzando un valor de US$4.248 billones en 2020[3]. En Estados Unidos, la penetración del *e-commerce* experimentó un crecimiento equivalente a diez años en apenas tres meses, mientras que las ventas minoristas sufrieron su mayor caída mensual (16.5% en abril de 2020)[4].

En respuesta a estas circunstancias, los negocios se vieron obligados a adaptarse rápidamente, buscando ofrecer experiencias mejoradas a un consumidor cada vez más exigente y en un mercado en constante expansión. Esto implicó realizar inversiones significativas en logística y cadenas de suministro. Además, los cambios en el entorno impactaron las decisiones de consumo; los usuarios se volcaron hacia nuevas marcas y alternativas, adaptándose a sus nuevas necesidades. En Estados Unidos, el 40% de los consumidores probaron nuevas marcas durante la pandemia, y un 46% dejó de ser leal a marcas conocidas[4].

Sin embargo, el principal *marketplace* de Latinoamérica, con 46 millones de compradores únicos en 2019[5], no enfrentó problemas significativos en los años siguientes. De hecho, la plataforma experimentó un crecimiento mayor al pre-pandémico, alcanzando 74 millones de compradores en 2022[5]. Pasó de un aumento anual del 10% en artículos vendidos antes de la pandemia a un crecimiento del 87% en 2020 y del 40% en 2021[5].

A medida que la oportunidad de mercado crecía, también se intensificaba la competencia con empresas locales, americanas y asiáticas, que invertían cada vez más en la región. En respuesta, la compañía en cuestión se enfocó en expandir y optimizar su red logística para hacerla más rápida, eficiente y rentable.

Entre otras mejoras, la empresa implementó tecnología y análisis de datos para mejorar la gestión de inventario en sus centros de almacenamiento en América Latina, que abarcan aproximadamente 1.700 millones de metros cuadrados. En estos centros, la empresa ofrece a los vendedores de la plataforma la opción de almacenar su inventario antes de la venta para acortar los tiempos de entrega, aumentar la conversión del sitio y minimizar los costos operativos mediante envíos multi-producto. Gracias a esta infraestructura, el 80% de los paquetes se entregan dentro de las 48 horas[5].

A pesar de los avances implementados para mejorar la rapidez de las entregas, la empresa aún enfrenta desafíos significativos en la toma de decisiones basadas en datos, especialmente en la predicción de demanda y la gestión eficiente del inventario. El principal dilema es determinar la cantidad óptima de *stock* de cada producto, un factor clave para calcular la capacidad necesaria en los centros de almacenamiento. Un exceso de *stock* es problemático debido a la limitación del espacio físico y el costo de oportunidad asociado, mientras que la escasez de *stock* puede resultar en pérdida de ventas y limitar la redistribución eficiente de productos entre almacenes.

Para abordar este desafío, la empresa utiliza algoritmos de *machine learning* para analizar las ventas de los últimos noventa días de cada SKU *(Stock Keeping Unit)* y proyectar la demanda futura. Sin embargo, la complejidad aumenta con productos nuevos sin historial de ventas, conocido como el problema de *cold-start.* En estos casos, la compañía limita a 20 unidades el *stock* inicial que un vendedor puede enviar por producto.

Como describen los autores del *paper*[6], "El problema de pronóstico de *cold start* ha planteado un desafío significativo en el campo de la predicción. Esta cuestión surge cuando no hay datos históricos disponibles para datos de series temporales o cuando los datos disponibles son insuficientes para hacer predicciones confiables". Este trabajo argumenta que, mediante técnicas de aprendizaje automático y el historial de productos de características similares, es posible optimizar esta decisión. Este estudio se enfoca en la vertical de electrónica de consumo, específicamente en celulares y computadoras, representando el 35% de la facturación mensual del servicio *fulfillment* de la compañía. El objetivo es

determinar si el umbral de 20 unidades es el más adecuado o si subestima o sobrestima la demanda potencial.

Con este fin, se entrenaron diferentes modelos de *machine learning* para predecir la demanda de productos nuevos, considerando atributos inherentes y patrones de otros productos similares. Utilizando técnicas de aprendizaje supervisado y diversos algoritmos, los resultados han sido prometedores. Un modelo destacó con un *F1-score* de 0.92 en datos de prueba, sugiriendo que ajustar el límite de unidades almacenadas podría reducir costos y aumentar ganancias, con una potencial disminución de hasta el 10% de unidades en los centros de distribución del *marketplace*.

1.1 Contexto

1.1.1 *¿Qué es E-commerce?*

El comercio electrónico, *o e-commerce*, refiere a la compra y venta de bienes o servicios a través de Internet, incluyendo la transferencia de dinero y datos necesarios para estas transacciones. Aunque comúnmente se asocia con la venta en línea de productos físicos, el *e-commerce* abarca cualquier tipo de transacción comercial realizada a través de Internet. Existen cuatro modelos principales:

1. *Business to Consumer (B2C)*: empresas venden a consumidores individuales (por ejemplo, una persona hace su compra de supermercado por la página *web*).
2. *Business to Business (B2B):* empresas venden a otras empresas (por ejemplo, una empresa que vende un servicio de *software* a otra empresa para su uso interno).
3. *Consumer to Consumer (C2C):* consumidores venden a otros consumidores (por ejemplo, cuando una persona vende su auto usado a un consumidor a través de una plataforma de compraventa).
4. *Consumer to Business (C2B):* consumidores venden a empresas (por ejemplo, un *influencer* que ofrece promocionar en sus redes los productos de cierta empresa a cambio de una tarifa).

El *e-commerce* está en constante evolución, con un enfoque creciente en el comercio móvil, la inteligencia artificial y la automatización para mejorar y personalizar la experiencia del cliente. En 2021, el comercio electrónico representó el 19% de las ventas minoristas globales, alcanzando aproximadamente 5 billones de dólares. Se espera que para 2026, abarque el 25% de las ventas minoristas globales, superando los 7 billones de dólares[7].

La madurez del *e-commerce* varía entre regiones, siendo Estados Unidos y Europa las más avanzadas, y Asia y América Latina las menos desarrolladas. Sin embargo, la pandemia de COVID-19 ha impulsado la expansión digital en regiones menos desarrolladas, con Brasil y Argentina liderando el crecimiento en *e-commerce*[8]. Este avance se relaciona con mejoras en el acceso a Internet, tecnología de dispositivos móviles, optimización logística y un mercado en expansión.

1.1.2 *Evolución del E-commerce*

La evolución del comercio electrónico ha avanzado desde las tradicionales tiendas físicas y los pedidos por correo hasta las actuales transacciones en línea. Las tiendas físicas, limitadas por costos de alquiler y necesidades operativas, contrastan con las compras por catálogo o por correo, precursores del *e-commerce*. En este último modelo, los vendedores anunciaban bienes y servicios en catálogos, revistas o televisión, y los clientes hacían pedidos por correo o teléfono, a menudo a través de un representante.

El origen de las compras en línea se remonta a los años 60 y 70, con la creación de redes como *ARPANET (Advanced Research Projects Agency Network)* que facilitaron la conectividad entre computadoras. En 1979, Michael Aldrich, un profesor de ciencias de la computación, inventó un sistema que vinculaba una televisión por cable a una computadora personal, conectándola a una tienda de electrodomésticos en el Reino Unido. Los clientes podían comprar productos a través de una línea telefónica, y se les cobraba en su factura telefónica. Este sistema, denominado "televenta" por Aldrich, es considerado uno de los primeros ejemplos de *e-commerce*[9].

Con los avances tecnológicos de los años 90, surgieron nuevos métodos de transacción electrónica que impulsaron el crecimiento del *e-commerce*. En 1994, First Virtual desarrolló un sistema de pagos electrónicos que permitía a los clientes pagar en línea con tarjeta de crédito[10]. Ese mismo año, Compaq Computer Corporation realizó la primera transacción comercial en línea de un producto físico, un CD de Sting, marcando un hito en la historia del comercio electrónico[9].

Uno de los primeros sitios de *e-commerce*, e incluso de los más importantes hasta el día de hoy, fue fundado en 1994 por Jeff Bezos: Amazon. La empresa comenzó vendiendo libros por la web ofreciendo reseñas de otros compradores para ayudar a los clientes a elegir, lo cual atrajo la atención de muchos usuarios. Rápidamente el sitio se fue expandiendo y pasó a ofrecer una amplia variedad de productos como CDs y DVDs entre otros. Un año más tarde Pierre Omidyar funda eBay, originalmente diseñada como una plataforma de subastas *online* donde los usuarios podían comprar y vender artículos de segunda

mano. Con el tiempo, la empresa fue ampliando su alcance para incluir también productos nuevos. Tanto Amazon como eBay son actualmente líderes en la industria, y sentaron las bases para el desarrollo de una gran cantidad de empresas de *e-commerce*[11].

El crecimiento exponencial del comercio electrónico en la década de 2000 fue impulsado por una mayor penetración de Internet y una creciente confianza en las compras en línea. Según un informe de la Conferencia de las Naciones Unidas sobre Comercio y Desarrollo (*UNCTAD*), las ventas de comercio electrónico en Estados Unidos crecieron de 5.5 mil millones de dólares a 8.5 mil millones de dólares entre 2012 y 2020[12]. Este crecimiento fue impulsado por tres factores principales:

1. Mayor adopción de la tecnología: El aumento en el uso de Internet y la mayor accesibilidad a dispositivos móviles y computadoras contribuyeron significativamente al incremento de las compras en línea. Un estudio de Pew Research Center revela que para el año 2000, el 52% de los adultos en Estados Unidos ya utilizaba Internet[13].
2. Mejoras en seguridad: a implementación de protocolos de seguridad como SSL (*Secure Socket Layer)* para proteger las transacciones en línea y la información personal fue crucial para aumentar la confianza de los consumidores en las compras en línea[14].
3. Desarrollo en logística: La aparición de compañías de logística especializadas y las mejoras en la infraestructura de envíos facilitaron entregas de productos más eficientes, lo que mejoró la experiencia de los usuarios. FedEx, por ejemplo, experimentó un crecimiento significativo en su división de *e-commerce* durante los primeros años de la década[15].

Estos factores han transformado los hábitos de compra, permitiendo a los consumidores acceder a una mayor variedad de productos y precios desde cualquier lugar y en cualquier momento. Este cambio ha tenido un impacto significativo en la economía digital, creando nuevas oportunidades de negocio y empleo. Actualmente, el comercio electrónico global tiene un valor de 7.60 billones de dólares y se proyecta que alcance los 16.24 billones de dólares para 2028[16]. VARStreet Inc[17] destaca la importancia de la inteligencia artificial y el aprendizaje automático en la evolución futura del *e-commerce,* anticipando una era de mayor personalización y eficiencia en las transacciones en línea.

1.1.3 *El Auge de los Marketplaces*

El término "*marketplace*", según el diccionario de Oxford[18], se refiere originalmente a "una plaza o lugar abierto en una ciudad donde se realizan mercados o ventas públicas". En el ámbito del comercio

electrónico, un *marketplace* se define como una plataforma digital que facilita la conexión entre compradores y vendedores, permitiendo transacciones de bienes o servicios[19].

Este concepto de *marketplace* tiene una larga historia, remontándose a la antigua Babilonia, donde se comerciaba con alimentos, ropa y cerámica[20]. Durante la Edad Media, estos mercados se formalizaron, organizándose en días específicos para ferias, como las famosas ferias de Champaña[21]. El surgimiento del dinero como medio de intercambio, alrededor del 1500 a.C. en Mesopotamia, jugó un papel crucial en la facilitación del comercio[22]. Con la Revolución Industrial desde 1760, la urbanización y la expansión de las rutas comerciales llevaron a la creación de mercados más grandes y permanentes. Innovaciones como el telégrafo y el ferrocarril impulsaron aún más el comercio[23]. Sin embargo, fue el surgimiento de Internet a fines del siglo XX lo que transformó radicalmente estos mercados, con sitios como Amazon[24] y eBay[25] liderando esta revolución digital.

En los últimos años, los *marketplaces* se han vuelto esenciales para el comercio electrónico, impulsados por la popularidad de las compras en línea, la economía colaborativa y avances tecnológicos que aseguran transacciones seguras. Estas plataformas ofrecen ventajas como la interacción directa entre compradores y vendedores, una variedad de productos y servicios, y herramientas adicionales que incluyen procesamiento de pagos, logística de envíos, *marketing* y publicidad. Los sistemas de calificación y opiniones, así como las opciones de mensajería y notificaciones, son fundamentales para mejorar la confianza y la calidad del servicio.

Según Vtex[26], los *marketplaces* se clasifican en horizontales y verticales. Un *marketplace* horizontal, como Amazon, busca ser un *One-Stop-Shop*, ofreciendo una variedad de productos y servicios de diferentes vendedores en un mismo lugar, facilitando una compra eficiente. En cambio, un *marketplace* vertical, como Etsy o Airbnb, se especializa en una categoría o industria específica, enfocándose en un nicho de mercado y ofreciendo productos especializados.

Logistics Brew[27] señala que los *marketplaces* generan ingresos de tres maneras: modelos de suscripción, donde se cobra una tarifa mensual a los vendedores; modelos de comisión, donde compradores y vendedores acceden gratuitamente y pagan solo cuando realizan transacciones; y cargos por publicar, donde se cobra al vendedor por listar productos. Entre los servicios que ofrece un *markeplace* se destacan:

1. Herramientas para publicar y motor de búsqueda: permiten a los vendedores publicar sus productos o servicios en la plataforma y ofrecen a los compradores herramientas para filtrar y ordenar listados según sus preferencias.
2. Sistema de calificaciones y opiniones: la plataforma permite a los consumidores evaluar su experiencia con el producto y el vendedor, lo cual es útil para otros compradores y motiva a los vendedores a mejorar la calidad de sus ofertas.
3. Mensajería y notificaciones: las plataformas proporcionan canales de comunicación, como correos electrónicos, SMS y alertas, para facilitar la interacción adecuada entre vendedores y compradores.
4. Sistemas de procesamiento de pagos: facilitan a los consumidores el pago de sus compras y a los vendedores la recepción del pago de manera segura, además de gestionar reembolsos y ofrecer financiamiento para incrementar las ventas.
5. Servicio al cliente: los *marketplaces* suelen ofrecer soporte tanto a compradores como a vendedores para resolver problemas y asegurar una experiencia positiva para todos.
6. *Marketing* y publicidad: incluyen listados priorizados en el motor de búsqueda o publicidad dirigida, permitiendo a los vendedores promocionar sus productos a una audiencia más amplia. Además, se ofrecen recomendaciones personalizadas basadas en el historial de navegación y compras del usuario.
7. Datos y análisis: proporcionan servicios de datos y análisis a compradores y vendedores para ayudarles a tomar decisiones informadas y mejorar su experiencia en la plataforma, incluyendo información sobre tendencias de ventas, comportamiento del cliente y rendimiento del producto.
8. Envíos y logística: muchas plataformas ofrecen servicios para gestionar la entrega de productos a los compradores, incluyendo integraciones para obtener información sobre el estado del envío, protección de la entrega y gestión de devoluciones.

1.1.4 *Servicios de Fulfillment Ofrecidos por los Marketplaces*

El auge del comercio electrónico ha transformado las expectativas de compra. Los consumidores, más exigentes que nunca, no solo esperan recibir sus productos rápidamente, sino que la velocidad de entrega a menudo determina su decisión de compra. Un estudio de McKinsey[28] revela que el 46% de los

consumidores han abandonado un carrito de compras debido a plazos de entrega prolongados o no especificados, mientras que un 34% opta por tiendas físicas por su inmediatez.

Además, no solo importa la rapidez de entrega. Según McKinsey[28], los consumidores priorizan primero los costos de envío (27.4%), luego la velocidad de entrega (26.6%) y finalmente la conveniencia de entrega en cuanto a ubicación y horario (23%). El cliente moderno busca productos con entregas rápidas y económicas, que combinan la comodidad de las compras online con la inmediatez de las tiendas físicas.

Este cambio en las expectativas se intensificó con Amazon Prime[29] en 2005, un programa de membresía que por $14.99 dólares al mes los consumidores pueden obtener envíos gratuitos en dos días. Este hito, estableció un nuevo estándar en la industria con envíos rápidos y gratuitos, alcanzando una rápida adopción en los consumidores. A abril de 2021, Amazon Prime reportó más de 200 millones de miembros globalmente[30].

En este contexto, los *marketplaces* asumen un papel crucial ayudando a los vendedores a satisfacer esta demanda creciente. Sus servicios de logística e infraestructura brindan soluciones integrales, permitiendo a los vendedores centrarse en el desarrollo y comercialización de sus productos, mientras la plataforma gestiona las entregas.

Esta alianza ofrece ventajas mutuas. Por un lado, los *marketplaces* negocian tarifas de envío más bajas debido al volumen de ventas, reduciendo el costo por paquete para vendedores y compradores. Por otro lado, al delegar el almacenamiento y empaquetado de productos, los vendedores pueden escalar sus operaciones, reduciendo costos y mejorando la rentabilidad sin necesidad de almacenes propios. Además, optimizando la cadena de envío, se minimizan los tiempos de entrega, mejorando la experiencia de compra e incrementando la conversión del sitio. Metapack[31] indica que la tasa de conversión puede aumentar hasta un 38% ofreciendo opciones de entrega adecuadas.

Dadas las razones expuestas, con la finalidad de aportar valor a compradores y vendedores, un análisis de la industria muestra que los *marketplaces* disponen de estructuras de red de variada complejidad, brindando una diversidad de servicios. La Figura 1.1 ilustra cómo un *marketplace* integra tres servicios logísticos distintos*: drop shipping, cross docking* y *fulfillment.*

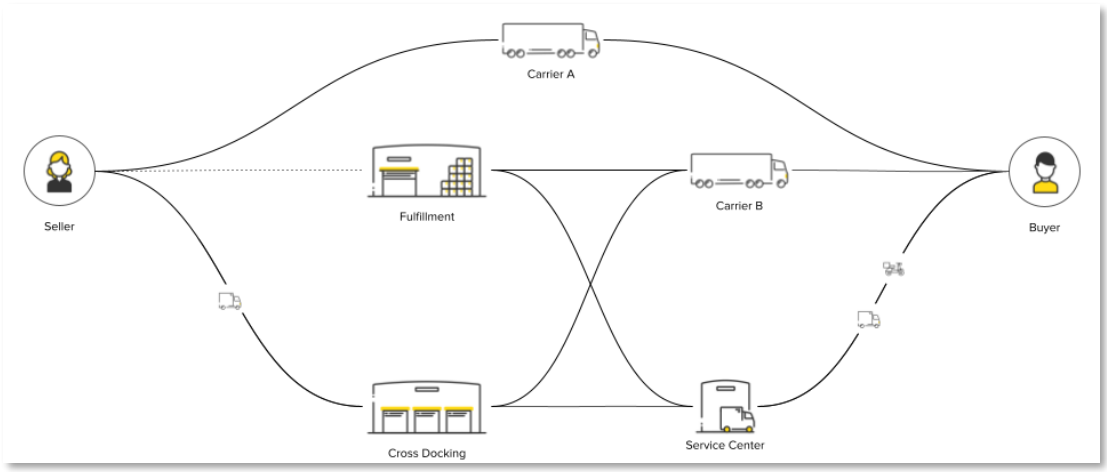


Figura 1.1: Diagrama de los tipos de servicio logístico ofrecidos por un marketplace

1. *Drop-Shipping*: este modelo logístico ofrece dos opciones principales: el vendedor puede almacenar su propio inventario o utilizar un almacén externo. Independientemente de la elección, tras una compra en el *marketplace*, el responsable (vendedor o tercero) debe preparar y empaquetar los productos, enviándolos a un punto de *drop-off* especificado por el *marketplace*. A continuación, una empresa de mensajería contratada por la plataforma recoge y entrega los paquetes al cliente final, utilizando su propia red logística para las millas intermedias. Esto, ilustrado en la Figura 1.2, muestra que el envío se hace directamente desde el vendedor hasta el comprador. Aunque este método mejora el control sobre la personalización de envíos y gestión de inventario, si el vendedor no utiliza almacenamiento externo, debe contar con su propia infraestructura, incrementando los costos de recursos, tiempo y esfuerzo laboral[32].

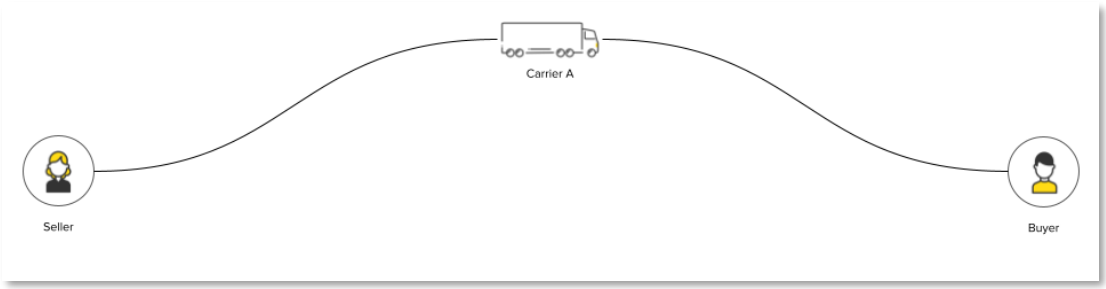


Figura 1.2 Modelo logístico de Drop-Shipping

2. *Cross-Docking:* como describe Shipbob[33], el *cross-docking* es una estrategia logística que optimiza la cadena de suministro al permitir la transferencia directa y eficiente de paquetes desde los vendedores hasta los clientes, sin necesidad de almacenamiento intermedio por parte de los *marketplaces*. Como se observa en la Figura 1.3, este proceso comienza cuando

se realiza una venta; el *marketplace* entonces moviliza su logística para recoger los paquetes de los vendedores, gestionando así la primera milla. Posteriormente, estos paquetes se trasladan a un nodo de *cross-docking*, donde se consolidan y se preparan para su próxima etapa: pueden ser enviados a un *service center* más cercano al cliente (representando la media milla) o directamente encaminados para su entrega final (última milla). Este enfoque no solo mejora la eficiencia en las rutas de entrega, sino que también reduce los costos de transporte al acercar los productos al consumidor final, beneficiando así tanto a vendedores como compradores con un servicio de mayor calidad.

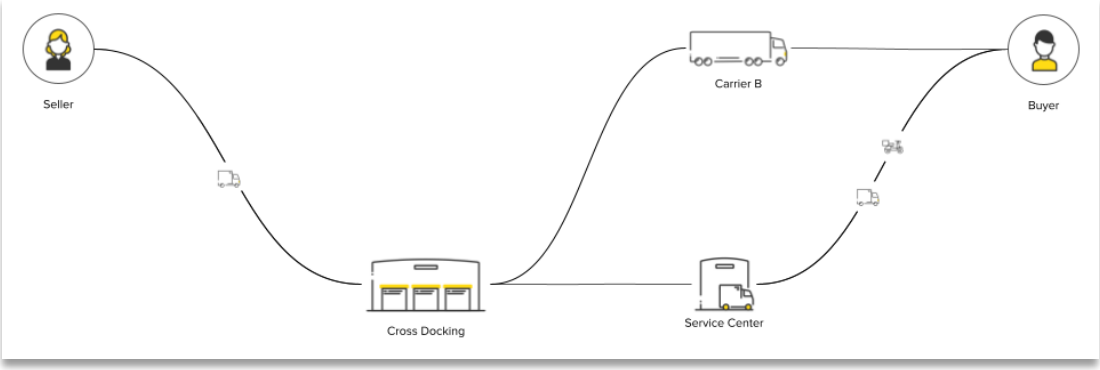


Figura 1.3 Modelo logístico de cross-docking

3. *Fulfillment by marketplace*: según describe Amazon[34], este programa ofrece a las pequeñas y medianas empresas los beneficios de su infraestructura logística al vender en la plataforma. Al seleccionar este modelo, los vendedores delegan la gestión de su inventario a los centros de distribución del *marketplace*. Estos, a su vez, coordinan la entrega de los productos optimizando la distribución a través de su red logística. Dependiendo de la ubicación del pedido, la plataforma decide el centro más conveniente desde el cual enviar el producto, asegurando una entrega eficaz ya sea directamente al cliente final o a través de nodos intermedios para acercar el producto (ver Figura 1.4). El servicio engloba desde la preparación del pedido hasta la atención al cliente postventa, incluyendo la gestión de posibles devoluciones.

   Este enfoque garantiza la calidad del servicio desde el embalaje hasta la entrega, mejorando significativamente la experiencia del usuario. Es especialmente ventajoso para pedidos que combinan productos de varios vendedores, ya que permite consolidarlos en un único envío, mejorando la eficiencia y reduciendo costos. La implementación de tecnología avanzada y

una infraestructura logística multi-nodo posibilita una rápida adaptación a las necesidades del consumidor, optimizando los tiempos de procesamiento y entrega.

Sin embargo, el éxito del modelo de *fulfillment by marketplace* depende fundamentalmente de una buena gestión de inventario; desde generar un pronóstico de ventas confiable, un minucioso seguimiento del nivel de *stock*, hasta una buena planificación de la reposición y su adecuado almacenamiento. Esto implica un trabajo en conjunto entre la plataforma y los vendedores para administrar el inventario de manera efectiva para garantizar que los productos no quiebren *stock*, que las ventas se procesen eficientemente y que los clientes reciban sus pedidos en el menor tiempo posible y a un precio competitivo. En este contexto, la plataforma es responsable de garantizar herramientas de gestión que brinden información precisa y actualizada sobre los niveles de inventario, tendencia de ventas y volúmenes de *stock* a reponer, para que los vendedores puedan garantizar la disponibilidad de existencias en plazos de reposición razonables.

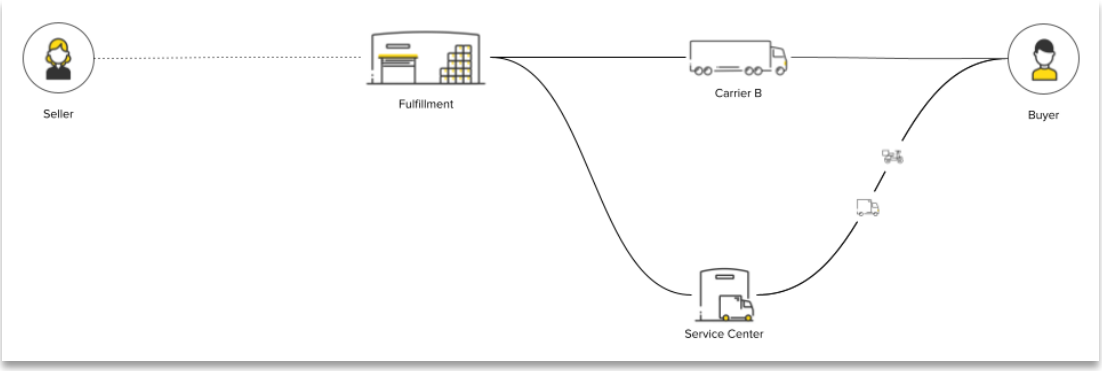


Figura 1.4 Modelo logístico fulfilment by marketplace

1.2 Problema

Supply Chain Management (SCM) se centra en garantizar un flujo eficiente de productos, servicios e información desde su punto de origen hasta que llegan a manos del consumidor. Este proceso está constantemente influenciado por cambios en elementos clave como la capacidad, la demanda del mercado y los costos. Estas fluctuaciones, particularmente en lo que respecta a la demanda, plantean desafíos significativos para la planificación y el manejo del inventario. Por ello, realizar proyecciones precisas de demanda se convierte en una herramienta clave para enfrentar estas incertidumbres, permitiendo una gestión más efectiva del inventario[35].

Para el *marketplace* líder en Latinoamérica, acertar en la predicción de la demanda resulta fundamental, sobre todo para su servicio de *fulfillment* para vendedores. Este reto se amplifica por distintos factores[36].

La primera complicación viene de la variabilidad en los patrones de consumo, que pueden alterarse drásticamente debido a nuevas tendencias, estacionalidades, promociones y eventos especiales, haciendo que la demanda sea impredecible.

En segundo lugar, la diversidad de productos en la plataforma, que incluye desde electrónica hasta moda y artículos para el hogar, introduce una complejidad adicional en la estimación de la demanda[37].

Un tercer aspecto es la competencia y la fluctuación de precios dentro del *marketplace*, donde la intensa rivalidad entre vendedores de productos similares puede causar volatilidad en la demanda, un fenómeno difícil de prever para la plataforma.

Los factores externos, como eventos económicos, sociales y climáticos, representan el cuarto desafío, ya que tienen el potencial de modificar radicalmente los patrones de consumo, quedando fuera del control directo del *e-commerce.*

En quinto lugar, el comportamiento omnicanal añade una capa más de complejidad, ya que la demanda de un producto puede variar significativamente entre diferentes canales de venta, complicando aún más las proyecciones.

Finalmente, la ausencia de datos históricos de ventas representa un desafío significativo para la estimación de la demanda. Este desafío, conocido como el "*cold-start*", alude a la dificultad de pronosticar la demanda sin información de ventas pasadas, similar a cómo un automóvil requiere alcanzar una temperatura óptima antes de operar eficientemente. Este dilema es el foco principal de esta tesis, dada su relevancia en el ámbito del comercio electrónico, donde la introducción constante de nuevos productos es la norma.

En el contexto de los grandes avances en *machine learning* de los últimos años, estas tecnologías se presentan como una solución prometedora ante dicho desafío. Sin embargo, a pesar de su potencial, la principal plataforma de comercio electrónico de Latinoamérica aún no ha implementado estas técnicas abordar el problema de "*cold-start*". En la actualidad, el *marketplace* comienza a proyectar las ventas de un producto solo después de que este ha acumulado un historial de ventas, utilizando los datos de los

últimos noventa días para estimar la demanda futura. Este enfoque deja un vacío en la estrategia de gestión de inventario durante el período crítico inicial tras el lanzamiento de un producto.

La precisión en la predicción de la demanda es crucial en un marketplace con servicios de *fulfillment*, ya que influye directamente en la gestión de inventario y la eficiencia de las bodegas. Mantener un equilibrio entre el exceso de *stock*, que conlleva costos de oportunidad y limitaciones de espacio, y la escasez de stock, que puede causar pérdida de ventas y dificultades en la redistribución de productos, es esencial. Por lo tanto, una estimación precisa de la demanda de cada SKU es fundamental para proporcionar a los vendedores orientación sobre la cantidad óptima de *stock* a ingresar.

Para determinar la cantidad de *stock* que un vendedor puede enviar de un producto específico, el *marketplace* en cuestión actualmente emplea un cálculo simple que se basa en la diferencia entre el pronóstico de ventas para las próximas seis semanas y el *stock* disponible en el centro de distribución. Este cálculo se expresa como:

$$Unidades\ a\ reponer\ SKU_i = Forecast\ de\ ventas\ SKU_i(6\ semanas)\ - Stock\ disponible\ SKU_i$$

Donde el "*forecast* de ventas (6 semanas)" es la cantidad proyectada de unidades que se espera vender en las próximas seis semanas, según estimaciones del modelo de aprendizaje automático. El "*stock* disponible" refiere a las unidades que el producto tiene actualmente disponibles para la venta en el centro de distribución.

Sin embargo, para productos nuevos, en lugar de basarse en el pronóstico de ventas, el *marketplace* establece un número fijo de unidades que un vendedor puede almacenar, el cual varía dependiendo de la capacidad ociosa en los centros de distribución de cada país. Según nuestra investigación, en Argentina el límite es de 50 unidades, mientras que en México y Brasil es de 20 unidades. Este umbral se establece de manera arbitraria, sin tomar en cuenta atributos específicos del producto. Solo después de que el producto comienza a tener ventas, el *marketplace* ajusta este umbral fijo a una estimación basada en los algoritmos de *machine learning,* que analizan las ventas de los últimos noventa días para proyectar la demanda futura.

Este enfoque arbitrario causa fricciones con los vendedores en la plataforma, generando una alta incidencia de consultas al servicio al cliente. Los vendedores señalan que adaptarse a este modelo de logística para productos nuevos puede ser desafiante, ya que un porcentaje significativo experimenta quiebres de *stock* poco después de ser listados, lo que requiere reabastecimientos frecuentes y, por ende, incrementa los costos operativos.

1.3 Revisión de literatura

A continuación, se presenta una revisión de los diferentes enfoques de aprendizaje automático propuestos en la literatura para abordar el desafío de estimar la demanda de productos sin un historial de ventas, destacando sus diferencias, fortalezas y debilidades.

1.3.1 *Aprendizaje No Supervisado*

El aprendizaje no supervisado, aunque menos común en la predicción de demanda, puede ser útil para identificar patrones y segmentaciones en los datos. Este enfoque descubre estructuras ocultas y relaciones dentro de los datos sin etiquetas de salida predefinidas.

Clustering

El *clustering* permite agrupar ítems similares sin necesidad de etiquetas explícitas, útil para el problema de *cold-start* sin datos históricos de ventas.

Los autores de "*Data Mining – Concepts and Techniques"* [38] argumentan que la selección de características relevantes es fundamental para aplicar *clustering* de manera efectiva. Las características pueden incluir datos explícitos, como marca, línea y detalles técnicos, e implícitos, como patrones de comportamiento de la demanda y preferencias. Estas características son esenciales para definir perfiles de ítems que luego se agruparán en *clusters*.

Los autores[39] discuten varios algoritmos de *clustering* que pueden ser aplicados en el contexto de estimación de demanda. Por ejemplo, el algoritmo *k-means* es popular debido a su simplicidad y eficacia en la formación de *clusters* basados en características numéricas bien definidas, mientras que otros algoritmos como *DBSCAN* pueden ser útiles para identificar *clusters* de densidad variable en datos más complejos. Se enfatiza la importancia de evaluar la calidad de los *clusters* formados, asegurándose de que los miembros de cada uno sean similares en términos de las características seleccionadas.

Una vez determinados los *clusters*, cuando se presentan nuevos ítems *(cold start),* se calculan sus características relevantes según la metodología propuesta y se asignan a los *clusters* existentes según su similitud con los grupos ya definidos. Posteriormente, la demanda puede ser estimada basada en los *clusters* identificados.

El *clustering* puede descubrir estructuras ocultas y relaciones dentro de los datos sin etiquetas de salida predefinidas. Sin embargo, su uso en la predicción de demanda tiene limitaciones importantes. La

calidad de los *clusters* depende en gran medida de la elección de características y parámetros, lo que puede ser subjetivo y requerir un conocimiento profundo del dominio. Además, la interpretación y validación de los *clusters* pueden ser complicadas.

En el contexto de esta tesis, el *clustering* fue descartado debido a estas limitaciones. Dado que la predicción de demanda requiere una estimación precisa y cuantificable, los métodos de *clustering*, que se centran más en la agrupación que en la predicción cuantitativa directa, no se consideraron adecuados para el objetivo específico de estimar la demanda de productos sin historial de ventas del *marketplace* en cuestión.

1.3.2 *Aprendizaje Supervisado*

El aprendizaje supervisado entrena modelos con datos etiquetados para predecir o clasificar nuevos datos. Utilizando productos similares con historial de ventas, es útil para estimar demanda en situaciones de *cold-start*.

Regresión Lineal

Uno de los modelos más comunes y simples en la literatura para este contexto es la regresión lineal. En el libro *Pattern Recognition and Machine Learning*[40], Bishop discute los pasos para aplicar este modelo. Primero, se recopilan datos de productos similares con historial de ventas, incluyendo características del producto y cantidades vendidas. Luego, se entrena un modelo de regresión con estos datos etiquetados. La regresión lineal encuentra la mejor línea recta que minimiza la diferencia entre las predicciones y las demandas reales. Después del entrenamiento y validación, el modelo se utiliza para predecir la demanda de nuevos productos sin historial de ventas, alimentando sus características en el modelo entrenado.

Bishop[41] argumenta que una ventaja de la regresión lineal es su simplicidad, lo que facilita su interpretación. Sin embargo, esta simplicidad puede ser una limitación, ya que puede no capturar relaciones no lineales complejas entre características del producto y la demanda. Además, la precisión del modelo depende de la calidad y disponibilidad de los datos de productos similares para el entrenamiento.

En el contexto de esta tesis, la regresión lineal fue descartada porque los datos del *marketplace* presentan una alta complejidad y relaciones no lineales que este modelo no puede capturar adecuadamente. Por lo tanto, se requieren modelos más sofisticados que puedan manejar estas complejidades.

Redes Neuronales

Las redes neuronales son modelos inspirados en el funcionamiento del cerebro humano, diseñados para identificar patrones complejos en los datos. Según Goodfellow, Bengio y Courville[42], son particularmente adecuadas para tareas que requieren modelar interacciones complejas y no lineales, siendo útiles para la estimación de demanda en situaciones de *cold start*.

El primer paso es recopilar datos de productos similares con historial de ventas conocido, incluyendo atributos relevantes como el precio y características técnicas. Luego, se entrena una red neuronal utilizando estos datos etiquetados. Las redes neuronales, mediante sus múltiples capas y nodos, pueden aprender representaciones complejas de los datos, capturando relaciones no lineales entre las características del producto y la demanda.

Se pueden utilizar diferentes arquitecturas de redes neuronales según la casuística del *marketplace*. Las redes neuronales *feedforward* (FNN) son adecuadas para problemas de predicción con una relación directa entre las entradas y las salidas, mientras que las redes neuronales recurrentes (RNN) son útiles cuando las características del producto incluyen secuencias temporales o series de tiempo, permitiendo que la red recuerde información de pasos anteriores, beneficioso para patrones temporales en la demanda.

Después de entrenar y validar la red neuronal, esta se puede utilizar para predecir la demanda de un producto sin historial de ventas, alimentando las características del nuevo producto (precio, características técnicas, etc.) en la red neuronal entrenada, que genera una estimación de la demanda esperada.

Según los autores del libro *Deep Learning*[43], una de las principales ventajas de las redes neuronales es su capacidad para modelar relaciones no lineales complejas en los datos, mejorando la precisión de las predicciones de demanda. Además, pueden manejar grandes volúmenes de datos y aprender de manera incremental a medida que se incorporan más datos. Sin embargo, las redes neuronales requieren grandes cantidades de datos para entrenarse adecuadamente, son computacionalmente intensivas y requieren mayor capacidad de procesamiento en comparación con modelos más simples. La interpretación de los modelos de redes neuronales también puede ser más complicada debido a su naturaleza de caja negra.

En la presente tesis, se optó por no utilizar redes neuronales debido a estas limitaciones prácticas, especialmente considerando la disponibilidad de recursos y la necesidad de interpretabilidad de los modelos en el contexto del *marketplace*.

Modelos Ensamblados

Los modelos ensamblados son una técnica de aprendizaje automático que combina múltiples modelos simples, cada uno entrenado de forma independiente, para mejorar la precisión. Según Hastie, Tibshirani y Friedman[44], modelos como *Random Forest* y *Gradient Boosting* manejan la no linealidad y se adaptan a estructuras complejas, por lo cual resultan útiles en contextos de *cold-start*.

El proceso inicia con la recopilación de datos de productos con historial de ventas, incluyendo atributos relevantes. Luego, se entrena el modelo ensamblado con estos datos etiquetados. Una vez entrenado y validado, el modelo se usa para predecir la demanda de nuevos productos sin historial de ventas, utilizando las características del nuevo producto para generar una estimación de demanda.

En línea los autores del libro *The Elements of Statistical Learning*[45], los modelos *Random Forest* y *XGBoost* resultan adecuados y por lo tanto fueron seleccionados en esta tesis por varias razones. Primero, son capaces de capturar relaciones no lineales complejas entre variables, lo cual es crucial para estimar la demanda de productos sin historial previo. Además, al combinar múltiples modelos base, reducen el riesgo de sobreajuste y son más robustos frente a diferentes estructuras de datos. A diferencia de las redes neuronales, estos modelos son menos intensivos computacionalmente y más fáciles de interpretar, lo que facilita su implementación práctica en el entorno del *marketplace*. Además, Random Forest y XGBoost han demostrado en la literatura un excelente rendimiento en una variedad de tareas de predicción, ofreciendo una buena compensación entre precisión y eficiencia computacional.

1.4 Objetivo

El objetivo principal de esta tesis es el desarrollo de un modelo de *machine learning* dirigido a pronosticar la demanda de productos nuevos, tomando como referencia el desempeño inicial de productos similares. Este análisis se fundamenta en datos aportados por el *retailer* más importante de la región, concentrándose en las ventas históricas en Brasil y México de productos estandarizados, tales como teléfonos y computadoras. Se examinan variables específicas como la marca, el modelo, el precio y las especificaciones técnicas, además del impacto que tiene la reputación del vendedor y su historial de ventas en la demanda de los productos nuevos. La meta es identificar patrones y tendencias que permitan afinar las proyecciones de demanda.

Esta iniciativa busca reemplazar la práctica actual del *marketplace*, que define de manera arbitraria la cantidad de inventario permitido para productos nuevos en *fulfillment*, por un método que

se base en el análisis de datos para estimar la demanda. De esta manera, se podrían tomar decisiones de gestión de inventario más informadas, ajustando adecuadamente los niveles de *stock* para prevenir tanto el excedente como la insuficiencia de inventario de productos nuevos en los centros de distribución, lo que conllevaría a una gestión más eficaz del inventario y a la minimización de costos relacionados con estimaciones incorrectas de la demanda.

Además, el modelo podría arrojar luz sobre las consecuencias de políticas restrictivas actuales, como el límite de 20 unidades por *SKU* para productos nuevos, y cómo estas pueden estar obstaculizando la introducción exitosa de dichos productos en el mercado. Si se valida la hipótesis de que tal restricción afecta adversamente a algunos segmentos, los resultados podrían motivar un ajuste en la política de inventario hacia una mayor flexibilidad. Esto no solo mejoraría el proceso de lanzamiento de productos en el servicio de *fulfillment*, sino que también beneficiaría tanto a vendedores como a compradores, optimizando tiempos de entrega y reduciendo costos de envío y, por ende, elevando la eficiencia logística y el nivel de servicio del *marketplace*.

Para lograr este objetivo, se exploraron modelos ensamblados como *Random Forest* y *XGBoost* en dos modalidades: regresión y clasificación. Estos modelos fueron seleccionados debido a su capacidad para manejar relaciones no lineales y su robustez frente a diferentes estructuras de datos, lo cual es crucial para estimar la demanda de productos sin historial previo. A diferencia de otros métodos considerados (*clustering*, regresión lineal y redes neuronales), los modelos de *Random Forest* y *XGBoost* ofrecen una combinación óptima de precisión, eficiencia computacional e interpretabilidad, haciendo posible una implementación práctica y efectiva en el entorno del marketplace.

2. Datos

La sección detalla el uso de datasets del mayor retailer de Latinoamérica para predecir la demanda de productos nuevos, describiendo la construcción del conjunto de datos, las categorías de características consideradas y un análisis exploratorio de variables clave, resaltando patrones y tendencias significativas.

Los datos utilizados para el desarrollo del modelo fueron proporcionados por el *retailer* más importante de América Latina, bajo un acuerdo de confidencialidad. A continuación, se especifica el uso de un conjunto seleccionado de *datasets*, cada uno con un propósito definido para la investigación:

- ***Dataset* de ítems*:*** cada registro pertenece a la publicación de un vendedor en particular, y se identifica unívocamente por la combinación de *item_id* y *variation_id*. Este *dataset* permite analizar las características de cada producto.
- ***Dataset* de actividad del ítem:** contiene registros de cada cambio realizado en las publicaciones a nivel *item_id* y *variation_id*, permitiendo reconstruir las condiciones y características de venta de los productos a lo largo del tiempo.
- ***Dataset* de reputación del vendedor:** información diaria de la reputación de los vendedores en la plataforma, incluyendo niveles de reclamos, cancelaciones y demoras en el despacho.
- ***Dataset* de reputación del ítem:** ofrece datos sobre la reputación diaria de cada producto, incluyendo reclamos y cancelaciones asociadas a cada *item_id.*
- ***Dataset* de ventas:** registra cada orden de compra (*order_id)*, por *item_id y variation_id*, incluyendo cantidad, precio y tipo de envío seleccionado por el consumidor.

El Apéndice 9.1 ofrece un listado completo de las variables disponibles en cada *dataset*, proporcionando una base detallada para el análisis realizado en esta tesis.

2.1 Estructura del *Dataset*

2.1.1 *Construcción del Target*

Para crear el conjunto de datos del modelo, se seleccionó información siguiendo criterios específicos. El proceso comienza con la selección cuidadosa de la información, empleando un enfoque de *snapshot* para capturar el estado exacto de las publicaciones de *notebooks* y celulares en Brasil y México desde el 1 de julio de 2021 hasta el 18 de mayo de 2022. Este método permite examinar las características de los productos en momentos clave: el día de la creación de la publicación y el día posterior. Así, para cada fecha dentro de este intervalo, se creó un registro único por producto, identificado por su *item_id* y *variation_id*, asegurando una captura detallada y precisa del inventario y sus cambios iniciales.

Se excluyeron del análisis los artículos usados o reacondicionados, aquellos productos que por su tamaño u otras características no podían ser enviados a través del *marketplace*, y los ítems cuya información no había sido verificada por el equipo de catálogo.

Para asegurar la precisión en la estimación de la demanda, también se excluyeron productos con menos del 50% de actividad en la plataforma durante las primeras seis semanas después de su publicación. Esto se determinó sumando los minutos de actividad desde su publicación hasta el término de las seis semanas, para evitar distorsiones causadas por periodos de inactividad. Por ejemplo, como se observa en la Figura 2.1, una publicación que estuvo activa 10.080 minutos en la semana 1 y 2, totaliza 20.160 minutos de actividad. Si este total no representa al menos el 50% del tiempo posible (calculado como 60 minutos * 24 horas * 7 días * 6 semanas), entonces el registro se excluye del *dataset*.

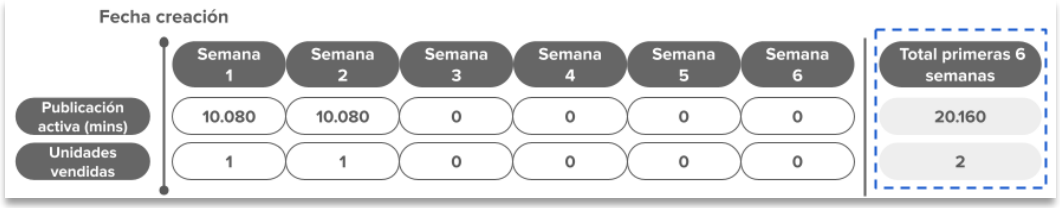


Figura 2.1 Ejemplo registros descartados por baja actividad en las primeras semanas de ventas

A continuación, se determinó el volumen de ventas de cada producto nuevo durante las primeras seis semanas tras su lanzamiento. Se sumaron las ventas semanales para obtener el total de ventas en este período, como se ilustra en la Figura 2.2.

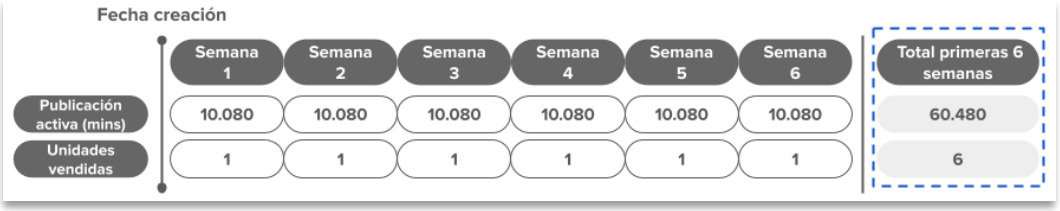


Figura 2.2 Ejemplo de cálculo de total de ventas por cada registro

Sin embargo, para una estimación precisa de la demanda utilizando el modelo, no se basó directamente en las ventas reales del producto en este lapso, sino en la demanda potencial que el producto habría alcanzado de haber estado disponible sin interrupciones durante las seis semanas completas. Por ejemplo, en la Figura 2.3 se muestra un producto que estuvo disponible un total de 40.320 minutos (67% del tiempo posible), vendiendo 4 unidades. Se asume que la inactividad en las semanas 5 y 6 pudo deberse a falta de *stock* o decisión del vendedor de pausar la publicación.

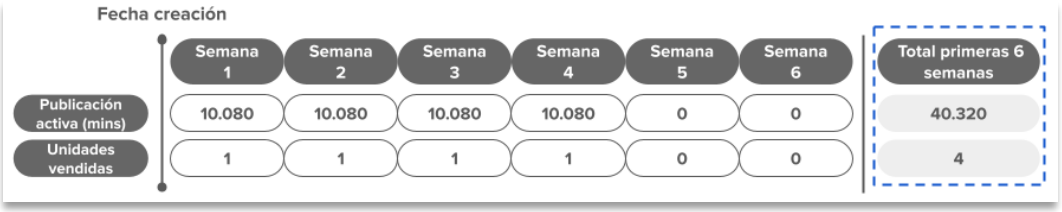


Figura 2.3 Ejemplo de cálculo de total de ventas por cada registro

Para estimar esta demanda potencial, se aplicó un método sencillo basado en una regla de tres, calculando las ventas esperadas si el producto hubiera estado activo durante los 60.480 minutos posibles de las seis semanas:

$$Demanda\ potencial\ (6W) = \frac{unidades\ vendidas\ (6W) * 60.480\ mins}{mins\ publicación\ activa\ (6W)}$$

En dicho caso, la demanda potencial para las primeras seis semanas sería de 6 unidades:

$$Demanda\ potencial\ (6W) = \frac{4\ unidades * 60.480\ mins}{40.320\ mins} = 6\ unidades$$

Este enfoque permite corregir el impacto de la gestión del vendedor sobre la disponibilidad del producto, aislando la demanda real del producto de las limitaciones de *stock*. Este cálculo de demanda potencial es el *target* que el modelo propuesto busca predecir.

La siguiente tabla proporciona ejemplos de cómo se aplicó este cálculo a registros específicos, mostrando la actividad, ventas reales y la demanda potencial calculada para productos en las primeras seis semanas desde su publicación:

snapshot	site_id	item_id	variation_id	minutes_active_N6W	units_sold_N6W	demanda_potencial_N6W
2021-10-16	MLB	2056282182	173810076969	60.480	42	42
2021-11-30	MLM	2100820898	173962910204	30.240	21	42

Figura 2.4 Vista simulada del conjunto de datos

2.1.2 *Construcción de Features*

La construcción de las características del modelo se estructuró en torno a tres ejes:

- ***Features* del vendedor:** La demanda de los productos está directamente influenciada por las características del vendedor. Un ejemplo claro es la reputación, un aspecto fundamental para los consumidores al tomar decisiones de compra. En el *marketplace* en cuestión, esta reputación se determina a partir de tres factores clave: tasa de reclamos, tasa de cancelaciones y rapidez de despacho. Estos elementos son cruciales para los compradores que buscan indicadores de confianza y calidad en el servicio, afectando así el posicionamiento de los productos en los listados de la plataforma.

- ***Features del ítem:*** se consideró información básica del ítem, como categoría y título, así como atributos clave para ofrecer a los clientes una visión completa del producto, tales como la cantidad de fotos y videos. También se incluyeron detalles relativos a las condiciones de venta, como el precio, disponibilidad de envíos gratuitos, opciones logísticas, retiros en tienda y garantías. Estos factores influyen significativamente en la elección del consumidor.

- ***Features* de la variante:** se incorporaron *features* específicos de cada variante del producto incluyendo marca, modelo, línea, código universal y características técnicas específicas como resolución de cámara, procesador y tamaño de pantalla. Estas características permiten distinguir las diferencias entre productos, facilitando a los consumidores una elección informada y ayudando a los vendedores a ajustar sus estrategias de precios y promociones.

Para manejar el dinamismo de las características y evitar el riesgo de filtración de datos (*data leakage*), se adoptó la estrategia de *snapshot*. Esta técnica consiste en tomar una fotografía de las características para cada registro en el momento de la creación de la publicación, proporcionando una base de datos estática y confiable para el entrenamiento del modelo.

En el Apéndice 9.2 se detallan todas las variables incluidas en el *dataset*.

2.2 Análisis Exploratorio de Datos

2.2.1 *Variable Dependiente: Demanda del SKU Nuevo*

La primera variable analizada es la que busca predecir el modelo: demanda del SKU nuevo en las primeras seis semanas de historia. Para lograrlo, se examinó el conjunto de datos completo, que comprende un total de 324.740 observaciones con valores desde 0 hasta un máximo de 5.205 ventas y una desviación estándar relativamente alta de 61,3.

La variable a estimar muestra una distribución altamente concentrada y sesgada hacia la izquierda, lo cual se refleja en la Figura 2.5.

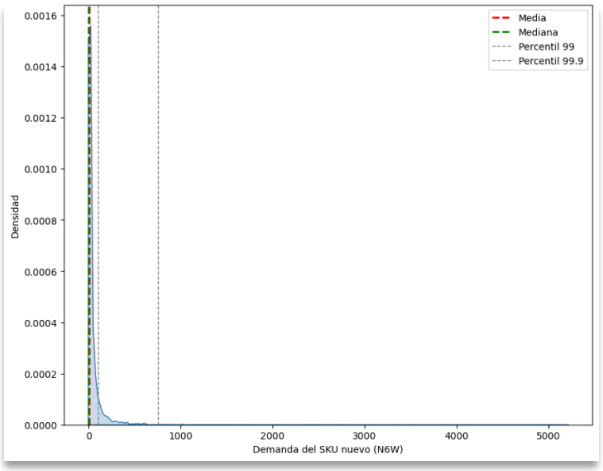


Figura 2.5 Distribución de unidades vendidas en las primeras 6 semanas

El análisis de la distribución hasta el percentil 95 de la Figura 2.6 a la izquierda, revela que la media de ventas es de 5,8 unidades, un valor bajo debido a la gran cantidad de observaciones con 0 ventas. La mediana, siendo también 0, resalta la concentración de datos en torno a esta cifra.

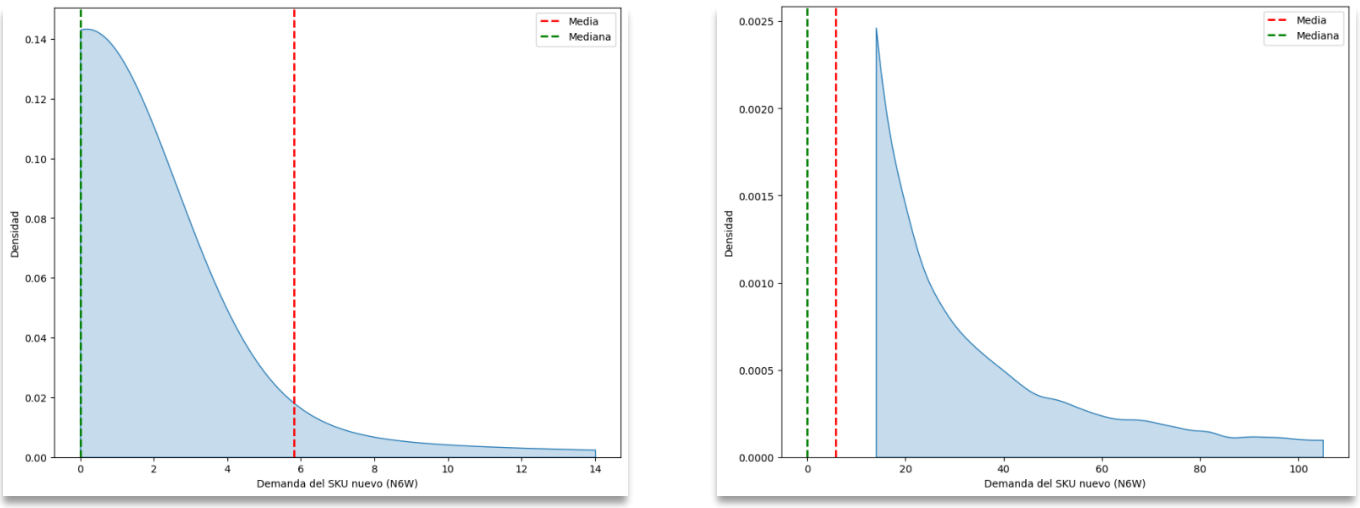


Figura 2.6 Izquierda: Distribución de unidades vendidas en las primeras 6 semanas (percentil 0 a percentil 95). Derecha: Distribución de unidades vendidas en las primeras 6 semanas (percentil 95 a percentil 99)

La Figura 2.6 a la derecha muestra la distribución de los percentiles 95 al 99, indicando que el 96% de los nuevos SKUs venden hasta 20 unidades en las primeras seis semanas. Esto sugiere que la regla del *marketplace*, que limita a 20 unidades la entrada de productos nuevos, es adecuada para la mayoría de los casos. Sin embargo, para el 4% de los SKUs que superan las 20 unidades vendidas en este período, la regla puede ser demasiado restrictiva. Esta limitación implica que los vendedores deban reabastecer el

producto con mayor frecuencia para evitar quebrar *stock*, resultando en costos de reposición más elevados o en la pérdida de ventas potenciales.

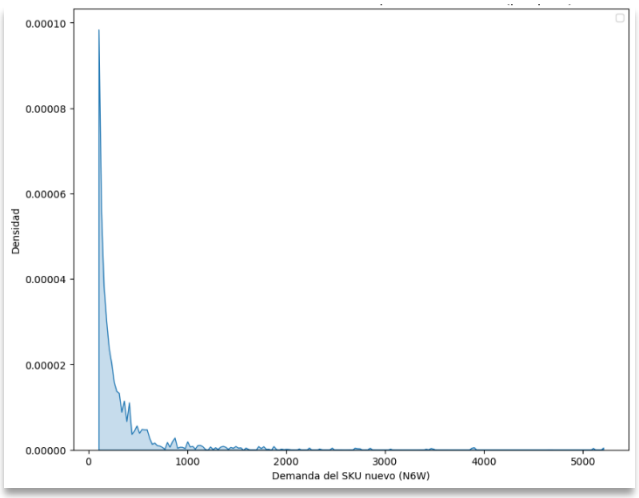


Figura 2.7 Distribución de unidades vendidas en las primeras 6 semanas (percentil 99 a percentil 100)

La Figura 2.7 muestra los *outliers* de la variable *target*, mostrando que el 1% de los SKUs supera las 105 unidades vendidas en las primeras seis semanas, con un máximo de hasta 5.205 unidades.

Es en este punto donde el modelo demuestra su valor, permitiendo identificar aquellos productos con un potencial de ventas superior. Esto abre la posibilidad de ajustar la regla de negocio actual, que limita a un máximo estático de unidades para *fulfillment*, para adaptarla a productos con alta demanda y, de esta manera, optimizar los ingresos de la plataforma.

2.2.2 *Valores Faltantes por Variable*

La Figura 2.8 muestra el porcentaje de valores faltantes para cada característica del conjunto de datos, segmentado en los tres grupos de variables: vendedor, ítem y variante.

Es relevante destacar que el conjunto de datos presenta una amplia variedad de características, siendo más numerosas para el ítem, seguidas por las de la variante y, finalmente, las del vendedor.

Del análisis se desprende que la única característica del vendedor con valores faltantes es *'nivel_reputacion_vendedor'*, la cual cuenta con información disponible para aproximadamente el 50% de las observaciones. Esta situación podría deberse a la presencia de vendedores nuevos en la plataforma que aún no han acumulado suficiente información para determinar su nivel de reputación.

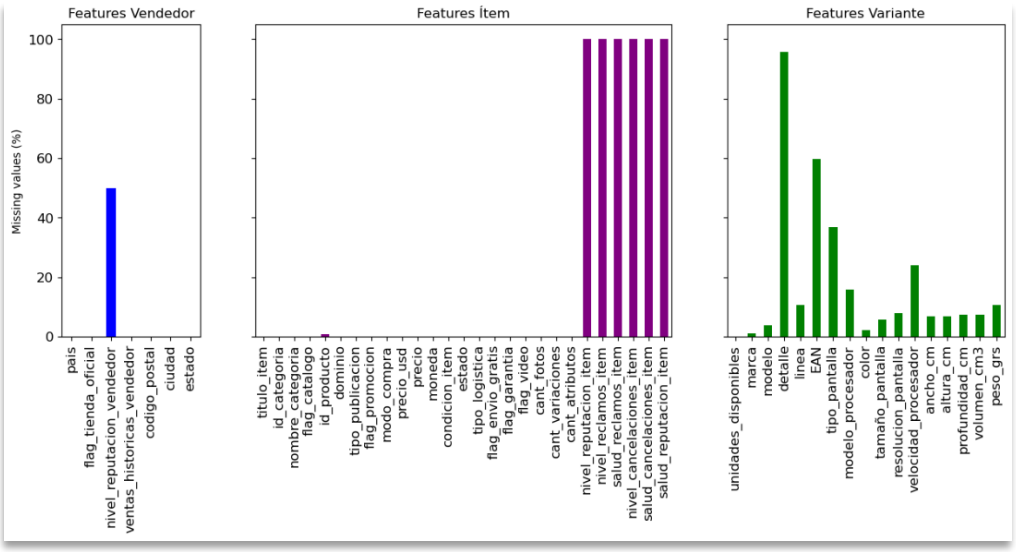


Figura 2.8 Valores faltantes por variable, segmentado por grupo

En relación con las variables asociadas a las características del ítem o publicación, se observa que generalmente no hay valores faltantes, con la excepción de aquellos relacionados con la reputación. Considerando que todas las observaciones en el conjunto de datos corresponden a productos nuevos, es lógico que no exista información sobre la reputación de estas publicaciones. Estas características no son pertinentes para nuestro análisis y, por lo tanto, pueden ser excluidas del modelo de entrenamiento.

En cuanto a las características relacionadas con la variante del producto, se destaca que la variable *'detalle'* tiene aproximadamente un 90% de valores faltantes, lo cual se explica porque es opcional para el vendedor incluir esta información al crear la publicación. Dado el enfoque del problema, esta variable no se considera relevante para el análisis. Por otro lado, la variable *'EAN'* presenta un 60% de observaciones faltantes. Aunque no siempre es proporcionada por el vendedor, el EAN es importante para identificar similitudes y detectar SKUs gemelos, lo que puede ser útil para el modelo al distinguir productos con características idénticas. Sin embargo, esta información puede ser complementada con otras variables como *'marca'*, *'modelo'* y *'línea'*. Las variables relacionadas con la pantalla (*'tipo_pantalla', 'tamaño_pantalla', 'resolucion_pantalla'*) y el procesador (*'modelo_procesador', 'velocidad_procesador'*) tienen entre un 10% y un 40% de valores faltantes. Además, las variables asociadas a las dimensiones y el peso del SKU presentan alrededor de un 10% de valores faltantes, pero como son características complementarias, no se espera que afecten significativamente el rendimiento del modelo.

2.2.3 *Correlación entre Features*

Con el propósito de explorar la relación entre las variables del conjunto de datos, se generó una matriz de correlación. Esta matriz proporciona una medida de la fuerza y dirección de la asociación lineal entre pares de variables. El coeficiente de correlación oscila entre -1 y +1, donde los valores cercanos a -1 indican una fuerte correlación negativa, los valores cercanos a +1 indican una fuerte correlación positiva y los valores cercanos a 0 indican poca o ninguna correlación. Una correlación positiva significa que un aumento en una variable se asocia con un aumento en la otra variable, mientras que una correlación negativa significa que un aumento en una variable se asocia con una disminución en la otra variable.

Es importante considerar que correlación no implica causalidad. Es decir, solo porque dos variables están fuertemente correlacionadas no significa que una variable cause la otra. Al interpretar una matriz de correlación, se pueden buscar patrones de asociación entre variables. Por ejemplo, si dos variables están fuertemente correlacionadas de manera positiva, puede indicar que están midiendo aspectos similares del mismo fenómeno. Por otro lado, si dos variables están fuertemente correlacionadas de manera negativa, puede indicar que están midiendo aspectos opuestos del mismo fenómeno.

A continuación, se examinan las correlaciones entre las características relevantes para el modelo con el objetivo de predecir la variable objetivo. La Figura 2.9 muestra una matriz de correlaciones positivas entre las variables numéricas del conjunto de datos. Aunque la mayoría de las correlaciones son moderadas (inferiores a 0,4), se destacan algunos patrones interesantes.

Es notable la correlación positiva entre *'demanda_N6W'* (demanda del nuevo SKU en las primeras seis semanas, la variable objetivo del modelo) y *'ventas_historicas_vendedor'* (corr: +0,22). Esta relación sugiere que los vendedores con un historial de ventas más alto tienden a tener un mejor desempeño en ventas para sus nuevos productos. Además, la conexión entre *'demanda_N6W'* y *'flag_tienda_oficial'* (corr: +0,17) es relevante, indicando que los productos nuevos de tiendas oficiales (vendedores que tienen autorización de la marca y de la plataforma para publicar ciertos productos) podrían tener mayores ventas en comparación con otros vendedores. Esto resalta la importancia de estas dos variables para el modelo, especialmente considerando que la presencia de una tienda oficial y un historial de ventas sólido se correlacionan positivamente (corr: +0,4), sugiriendo una sinergia en la influencia sobre las ventas de nuevos productos.

Las demás correlaciones identificadas son generalmente débiles, lo que indica que, aunque existen patrones interesantes, la mayoría de las variables numéricas tienen una influencia limitada sobre la variable objetivo de manera individual.

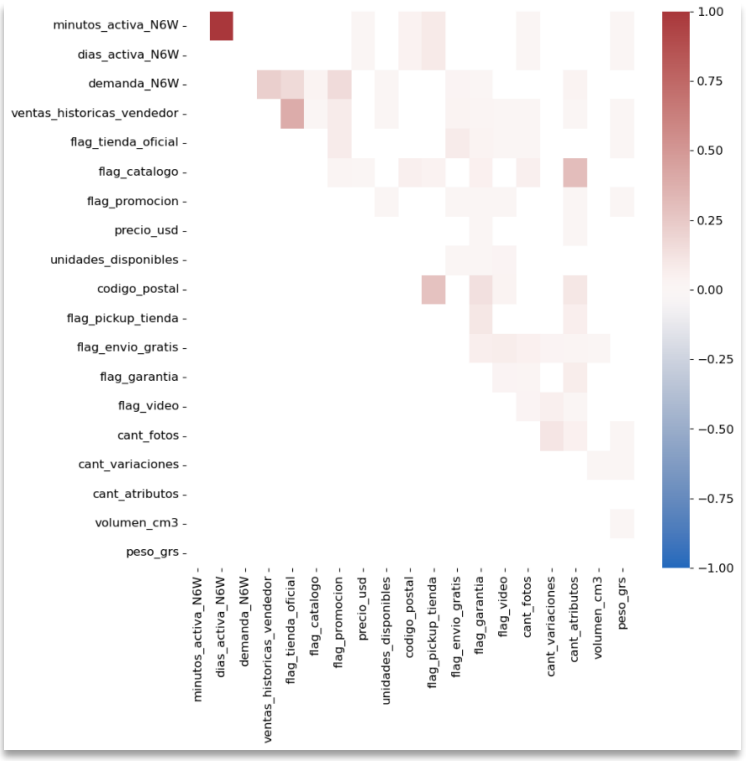


Figura 2.9 Matriz de correlaciones positivas

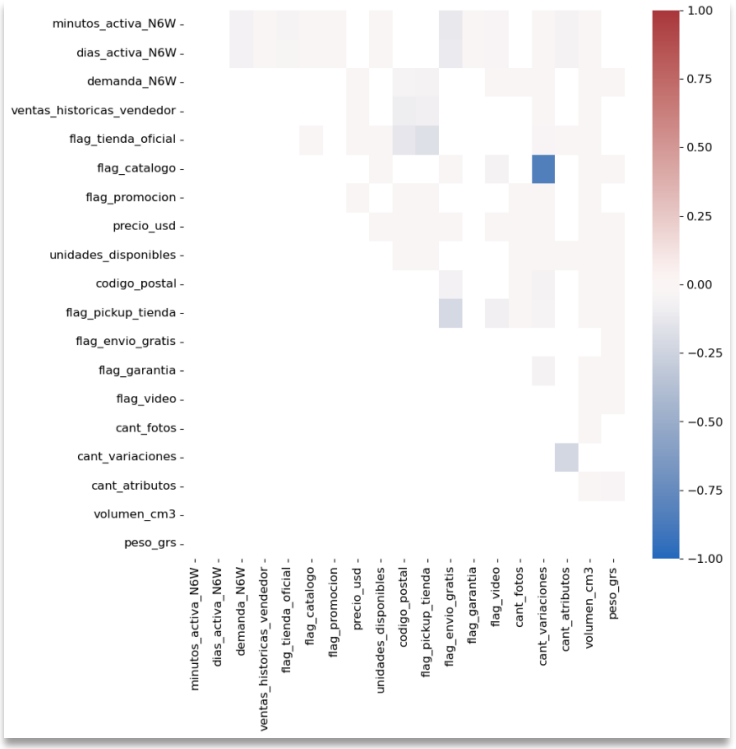


Figura 2.10 Matriz de correlaciones negativas

En la sección del Apéndice 9.3 se detallan todas las correlaciones entre *features*.

2.2.4 *Relación entre Features y Variable Dependiente*

2.2.4.1 Features del Vendedor

País del Vendedor

El conjunto de datos muestra un desbalance en la variable *'país'*, con un 80% de los registros correspondientes a productos de Brasil y el 20% restante a México. Esta desproporción en la distribución de los datos es un aspecto crucial para el modelado, ya que el modelo cuenta con menos información de México para detectar patrones y tendencias.

Al explorar la relación entre el '*país*' y la demanda del nuevo producto durante las primeras seis semanas, se observan diferencias significativas. En Brasil, un porcentaje más elevado de observaciones reporta cero ventas (85% en Brasil frente a 70% en México). Además, como se ilustra en la Figura 2.11, los volúmenes de venta para los SKU nuevos que logran vender son notablemente mayores en México. Específicamente, los SKU nuevos en México venden 2,8 veces más que en Brasil (el percentil 99 de ventas en Brasil es 59, mientras que en México es 167).

País	Media	Mediana	Perc'25	Perc'50	Perc'75	Perc'90	Perc'95	Perc'99
Brasil	2,7	0	0	0	0	3	8	59
México	7,9	0	0	0	2	15	40	167

Figura 2.11 Distribución de la demanda de SKU nuevo por país

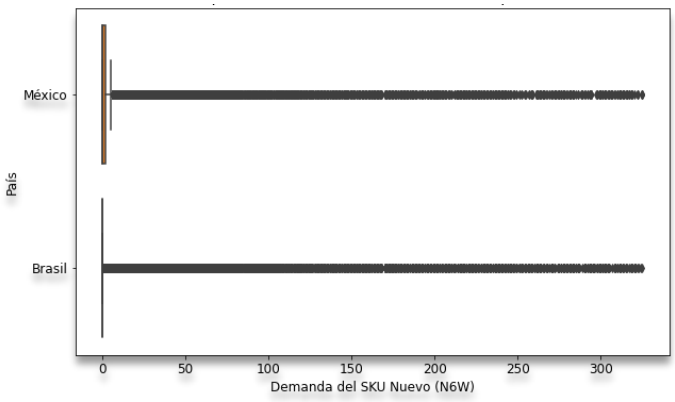


Figura 2.12 Boxplot de la demanda del SKU nuevo según país

Esta variación en el desempeño de ventas entre los países tiene implicaciones directas para el modelado del problema. Sugiere que el rendimiento de un producto nuevo puede variar según el país. Esto no solo indica que Brasil tiene una mayor proporción de SKUs sin ventas, sino que también resalta la capacidad de México para alcanzar volúmenes de venta más altos para los SKU que sí venden.

Para el *marketplace* en cuestión, la política de flexibilizar la cantidad de unidades permitidas en *fulfillment* debería considerar estas diferencias por país. La inclusión de esta variable en el modelado

permitirá una predicción más precisa del comportamiento de la demanda y una adaptación más eficiente de las estrategias según el contexto geográfico.

Flag Tienda Oficial

Las tiendas oficiales en la plataforma gozan de una visibilidad privilegiada y ocupan un lugar destacado en el sitio *web*, lo cual se refleja en su insignia distintiva. El estatus de tienda oficial es otorgado por el equipo comercial de la plataforma a negocios seleccionados.

A pesar de que solo el 7% de las observaciones en el conjunto de datos corresponden a vendedores catalogados como tienda oficial, este distintivo tiene un impacto considerable en la demanda. La Figura 2.13 ilustra diferencias estadísticamente significativas en la demanda de un SKU nuevo con el *flag* de tienda oficial.

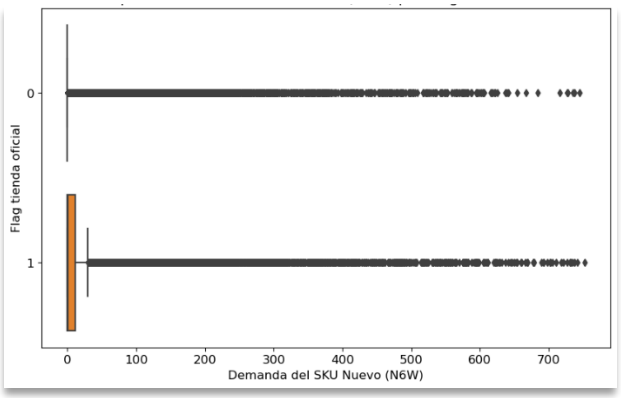


Figura 2.13 Boxplot de la demanda del SKU nuevo según Flag de tienda oficial

En promedio, un SKU de una tienda oficial experimenta una demanda 14 veces mayor en las primeras seis semanas que uno de una tienda no oficial (43 vs 3 unid.), como se ilustra en la Figura 2.14.

Tienda oficial	Media	Mediana	Perc'25	Perc'50	Perc'75	Perc'90	Perc'95	Perc'99
0	3	0	0	0	0	3	9	59
1	43	1	0	1	13	69	177	856

Figura 2.14 Distribución de la demanda de SKU nuevo por Flag de tienda oficial

Este análisis subraya la relevancia de la condición de tienda oficial para la demanda de un SKU nuevo, un factor clave en el modelado y la predicción. Además, tiene implicancias importantes para el *marketplace*, ya que indica que la variable de tienda oficial es un factor a considerar al definir qué política aplicar en sus *fulfillments* para permitir el ingreso de unidades cuando un SKU es nuevo.

Reputación del Vendedor

La reputación del vendedor es un indicador crucial en la plataforma, reflejando la calidad del servicio y la atención al cliente. Este índice se construye considerando aspectos como cantidad de reclamos, cancelaciones y retrasos en los despachos.

La Figura 2.15 muestra la distribución de la demanda de SKUs nuevos según los niveles de reputación de los vendedores. La mayoría de la demanda (70%) se concentra en vendedores de reputación *'green_platinum',* el nivel más alto, seguido por el 20% en vendedores de reputación *'green',* un nivel intermedio. Los demás niveles concentran menos del 5% cada uno.

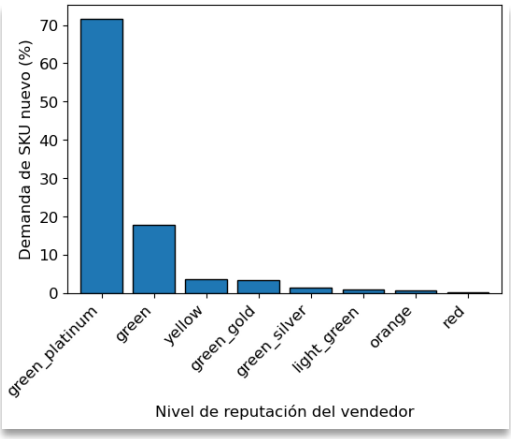


Figura 2.15 Demanda de SKU nuevo según reputación del vendedor

Nivel de reputación	Media	Mediana	Perc'25	Perc'50	Perc'75	Perc'90	Perc'95	Perc'99
green platinum	28,5	1	0	1	10	47	112	487
green gold	4,8	0	0	0	2	8	20	80
green silver	3	0	0	0	2	6	13	46
green	6,1	0	0	0	2	8	20	100
light green	3	0	0	0	0	3	6	31
yellow	3,9	0	0	0	2	6	15	62
orange	1,5	0	0	0	0	2	5	24
red	0,4	0	0	0	0	0	2	7

Figura 2.167 Distribución de la demanda de SKU nuevo por nivel de reputación del vendedor

La Figura 2.16 detalla la distribución de la demanda de SKUs nuevos según el nivel de reputación, destacando que los vendedores *'green_platinum'* tienen en promedio 28,5 ventas, significativamente más que el resto, cuyas ventas oscilan entre 0,4 y 6,1. Esto sugiere que una buena reputación puede aumentar significativamente la demanda de los productos vendidos en la plataforma.

Este hallazgo es relevante para el modelado del problema, ya que subraya la importancia de incorporar la reputación del vendedor como una variable predictiva clave en la estimación de la demanda de un SKU nuevo. Al tener en cuenta este factor, la plataforma puede desarrollar estrategias de gestión

de inventario más efectivas. En particular, puede implementar estrategias diferenciadas de ingreso de inventario para productos nuevos adaptándolas según la reputación de los vendedores para maximizar las ventas.

Histórico de Ventas del Vendedor

La relación entre la demanda de un SKU nuevo y el historial de ventas del vendedor se examina en la Figura 2.17. Los resultados no muestran una correlación directa y evidente entre las variables lo que sugiere que hay otros elementos que influyen en las ventas de un SKU nuevo más allá del historial del vendedor.

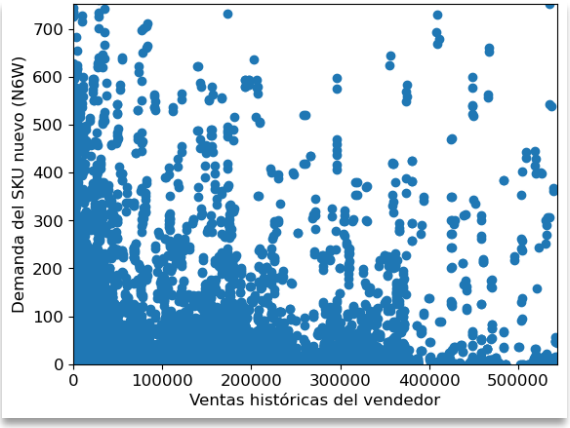


Figura 2.17 Demanda de SKU nuevo según ventas históricas del vendedor

Si bien el historial de ventas puede ser un indicador útil, no es el único factor que determina el éxito de un producto nuevo. Aspectos como la calidad del producto, la demanda del mercado, la competencia y la estrategia de marketing también juegan un papel importante en la determinación de las ventas. Por lo tanto, el modelo debe incluir no solo el historial de ventas, sino también factores para obtener una visión más completa y precisa del comportamiento de la demanda.

2.2.4.2 Features del Ítem

Dominio del Ítem

El dominio se define como la combinación entre la categoría del producto (celulares/*notebooks*) y el país donde fue publicado (MLB: Brasil y MLM: México). La Figura 2.18 muestra que más del 50% de la demanda de SKUs nuevos se concentra en celulares de Brasil, 40% en celulares de México y el resto se distribuye equitativamente entre *notebooks* de ambos países.

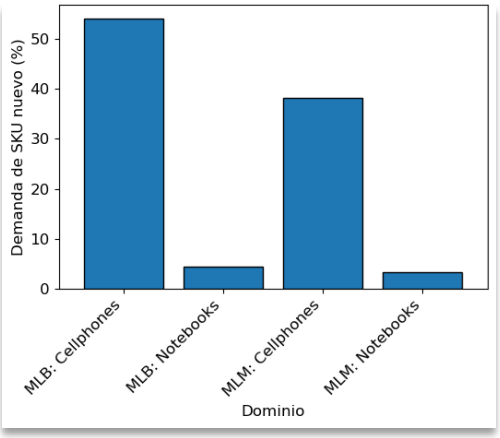


Figura 2.18 Demanda de SKU nuevo según dominio del ítem

Al analizar la relación entre la demanda del SKU nuevo y el dominio en la Figura 2.19, se puede destacar que el dominio de celulares en México tiene una demanda mucho más alta que el resto de los dominios. En promedio, las publicaciones de celulares en México venden 9,8 unidades en las primeras seis semanas, en comparación con un promedio de ventas de entre 2,4 y 2,7 unidades para los otros dominios.

Dominio	Media	Mediana	Perc'25	Perc'50	Perc'75	Perc'90	Perc'95	Perc'99
MLB: Cellphones	2,7	0	0	0	0	3	9	60
MLB: Notebooks	2,4	0	0	0	0	3	5	53
MLM:Cellphones	9,8	0	0	0	3	21	50	200
MLM:Notebooks	2,6	0	0	0	0	2	8	68

Figura 2.19 Distribución de la demanda de SKU nuevo por dominio del ítem

Este hallazgo es relevante para el modelado del problema, ya que sugiere que el dominio del ítem puede ser un factor predictivo importante para la demanda de un SKU nuevo. Considerar este factor permite al modelo identificar segmentos de mercado con potencial de ventas más alto, como los celulares en México, y permite al *marketplace* ajustar las políticas de ingreso de inventario de productos nuevos de manera más efectiva para maximizar las ventas en esos dominios.

Tipo de Publicación

La variable *'tipo_publicacion'* se clasifica en tres categorías: *'gold_pro'* (publicación *premium*), *'gold_special'* (publicación clásica) y *'free'* (publicación gratuita). Cada categoría ofrece distintas ventajas y restricciones, resumidas en la Figura 2.20.

	Free	*Gold Special*	*Gold Pro*
Costo por Publicar	$ 0	$ 0	$ 0
Exposición en los listados	Baja	Alta	Máxima
Duración	60 días	Ilimitada	Ilimitada
Ofrece cuotas sin interés	NO	NO	SI
Costo por vender *(según categoría del producto)*	$ 0	México: 8% - 16% Brasil: 10% - 14%	México: 12.5% - 20.5% Brasil: 15% - 19%

Figura 2.20 Características de cada tipo de publicación

La Figura 2.21 muestra que el tipo de publicación *'gold_pro'* concentra 55% de la demanda de productos nuevos, seguido de cerca por *'gold_special'* con el 45%. La categoría *'free'* capta una porción mínima de la demanda. Esto se debe al mejor posicionamiento en los listados que tienen las publicaciones pagas, aunque no siempre ofrezcan los precios más competitivos.

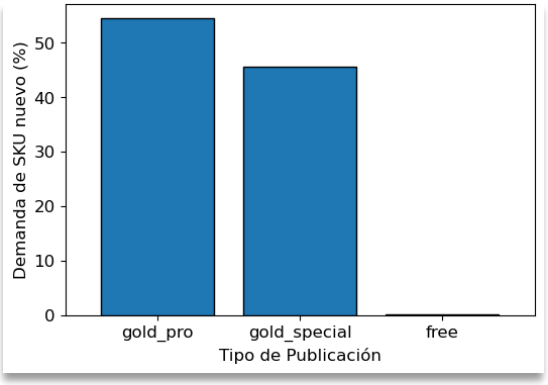


Figura 2.21 Demanda de SKU nuevo según tipo de publicación

La Figura 2.22 revela que 99% de las publicaciones *'free'* no registran ventas en las primeras seis semanas, indicando que tienen menor probabilidad de generar ventas en comparación con las pagas.

Tipo de publicación	Media	Mediana	Perc'25	Perc'50	Perc'75	Perc'90	Perc'95	Perc'99
Free	0	0	0	0	0	0	0	0
Gold Special	4,7	0	0	0	0	6	20	110
Gold Pro	9,8	0	0	0	1	6	18	100

Figura 2.22 Distribución de la demanda de SKU nuevo por tipo de publicación

Estos hallazgos son relevantes para el modelo ya que sugieren que el tipo de publicación puede ser un factor predictivo importante para estimar la demanda de un SKU nuevo. La preferencia de los compradores por publicaciones con mayor visibilidad y la disposición a pagar por características *premium* como cuotas sin interés podrían influir significativamente en las ventas de un SKU nuevo.

Flag Promoción

El *'flag_promocion'* indica si una publicación tenía una campaña activa en el momento de su creación. Aunque solo el 0,06% de las observaciones en el conjunto de datos tenían este *flag* activado, es relevante analizar su relación con la demanda en las primeras seis semanas de un producto nuevo.

Como se observa en la Figura 2.23, menos de 1% de los SKUs nuevos tenían el *flag* de promoción igual a 1 (activo) en al momento de creación de la publicación. Sin embargo, la Figura 2.24 muestra que SKUs con *flag* de promoción activa experimentaron una demanda promedio hasta 70 veces mayor que

aquellos sin promoción. Esto sugiere el impacto considerable de las campañas en la demanda de productos nuevos.

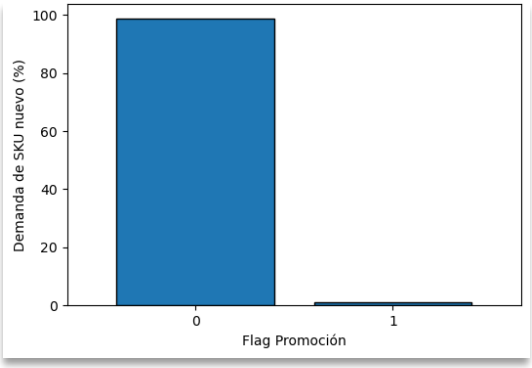


Figura 2.23 Boxplot de la demanda del SKU nuevo según Flag promoción

Flag promoción	Media	Mediana	Perc'25	Perc'50	Perc'75	Perc'90	Perc'95	Perc'99
0	5,5	0	0	0	0	4	14	103
1	394,3	46	10	46	302	1.135	2.346	3.596

Figura 2.24 Distribución de la demanda de SKU nuevo por Flag promoción

Este descubrimiento es clave para el modelo, ya que sugiere que la presencia de promociones puede ser un factor predictivo importante para estimar la demanda de un SKU nuevo. El *marketplace* podría incluso incentivar a los usuarios a utilizar más estas herramientas promocionales para impulsar la rotación de sus productos, ya que los resultados son muy favorables.

Precio del SKU (USD)

La relación entre precio y demanda de SKUs nuevos en las primeras seis semanas se presenta en la Figura 2.25. Se observa una tendencia en la que los productos con precios más bajos tienden a tener un mayor volumen de ventas, mientras que aquellos con precios más altos experimentan menores volúmenes de ventas. La mayoría de las observaciones se encuentran en el rango de 0 a 500 USD.

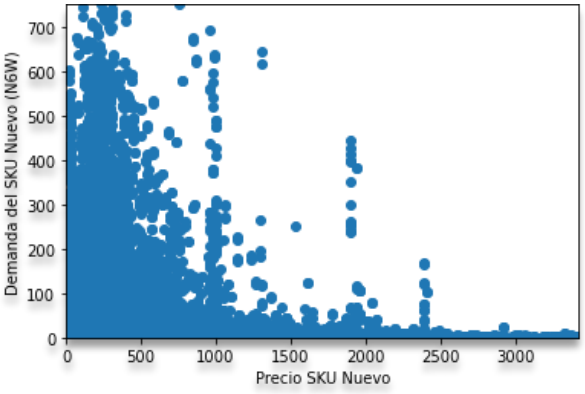


Figura 2.25 Demanda de SKU nuevo según precio (USD)

Es relevante mencionar que la presencia de precios extremadamente bajos en computadoras y celulares nuevos podría deberse a modelos antiguos o a valores atípicos en el conjunto de datos.

Al modelar la demanda de SKUs nuevos y definir cuántas unidades permitir almacenar, es importante considerar la relación precio-demanda para optimizar la gestión de inventario y maximizar las ventas en el *marketplace*.

Tipo de Logística

La variable *'logistic_type'* clasifica el método de envío seleccionado por el vendedor, con cuatro opciones disponibles:

1. *'drop_off':* el vendedor envía el producto por correo.
2. *'xd_drop_off'*: el vendedor lleva la venta a puntos de entrega autorizados, desde donde la plataforma gestiona las entregas.
3. *'cross_docking'*: la plataforma recolecta las ventas por el domicilio del vendedor y las envía al comprador.
4. *'fulfillment'*: los productos se almacenan en los depósitos de la plataforma, quien se encarga de gestionar toda la entrega.

Al analizar la relación entre el tipo de logística y la demanda del SKU nuevo en la Figura 2.26, se observan claras diferencias. Por ejemplo, el tipo de logística '*fullfilment*' muestra un volumen de ventas mucho mayor que el resto, con una media de 50 unidades demandadas por SKU en comparación con 12,3 en '*cross_docking'*, 6,2 en '*XD_drop_off*' y 1,5 en '*drop_off*'.

Tipo de logística	Media	Mediana	Perc'25	Perc'50	Perc'75	Perc'90	Perc'95	Perc'99
Drop off	1,5	0	0	0	0	2	4	32
XD drop off	6,2	0	0	0	2	11	27	128
Cross docking	12,3	0	0	0	5	28	63	224
Fulfillment	50	14	3	14	54	151	254	380

Figura 2.26 Distribución de la demanda de SKU nuevo por tipo de logística

La figura 2.27 refuerza este punto, donde se observa que hay diferencias estadísticamente significativas entre la demanda de un SKU nuevo que vende a través de *fulfillment* con relación a uno que ofrece otro tipo de logística.

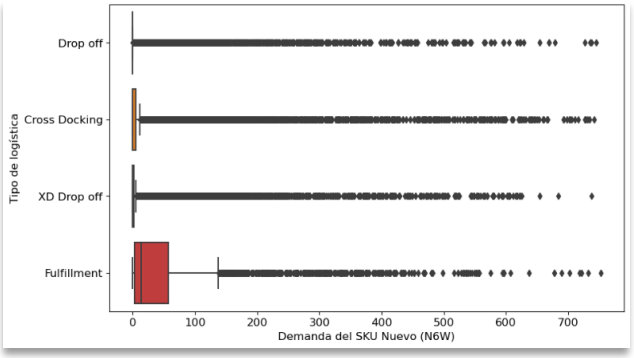


Figura 2.27 Boxplot de la demanda del SKU nuevo según tipo de logística

Estos resultados subrayan la relevancia del tipo de logística en la demanda de SKUs nuevos, sugiriendo que el marketplace debería promover el uso de *'fulfillment'* entre los vendedores para impulsar las ventas. Además, la relación entre la logística elegida y la demanda esperada es crucial para la estrategia del *marketplace* en cuanto a la gestión de inventario de SKUs nuevos. Es esencial que la plataforma adapte su política de *stock* máximo considerando esta relación, especialmente si un producto nuevo utiliza *'fulfillment'* como logística predeterminada, lo cual se espera que incremente notablemente las ventas y, por ende, debería reflejarse en la cantidad de unidades permitidas para almacenar.

Cantidad de Atributos de la Publicación

En este conjunto de datos, solo se incluyen publicaciones con información respaldada por el catálogo del *marketplace*. Esto significa que los atributos del producto se completan automáticamente al asociar la publicación con un producto existente en el catálogo, proporcionando a los compradores información detallada sobre el producto y facilitando una decisión de compra informada y sin fricciones.

La Figura 2.28 muestra que una publicación puede incluir hasta 88 atributos, aunque en promedio se incluyen alrededor de 66.

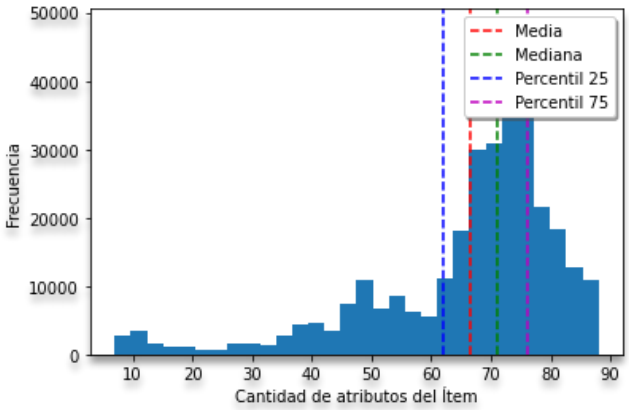


Figura 2.28 Distribución de observaciones según cantidad de atributos del ítem

La Figura 2.29 ilustra la relación entre la cantidad de atributos del ítem y la demanda durante las primeras seis semanas. Se observa una relación positiva entre las variables, a medida que aumenta la cantidad de información proporcionada en la publicación, también aumenta la demanda del producto.

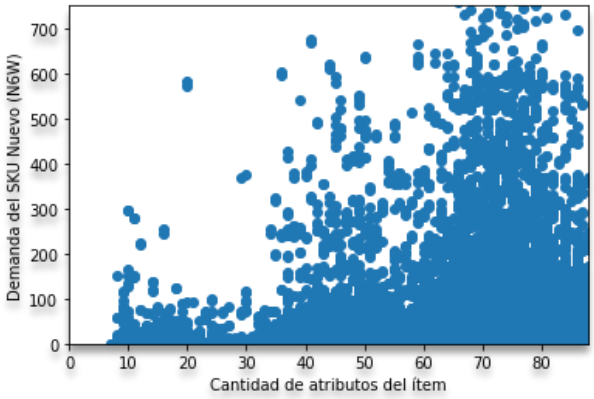


Figura 2.29 Distribución de la demanda según cantidad de atributos del ítem

Estos hallazgos son importantes para el modelado del problema, ya que indican que la cantidad de atributos es un factor relevante en la determinación de la demanda de un SKU nuevo. Incluso el *marketplace* podría tener diferentes políticas de ingreso de inventario en *fulfillment* según la calidad de la publicación en términos de cantidad de atributos, con el fin de maximizar las ventas.

2.2.4.3 Features de la Variante

Unidades Disponibles a la Venta

La Figura 2.30 a la izquierda analiza la cantidad de unidades disponibles a la venta en cada SKU al momento de la creación de la publicación. 53% de los ítems tienen una sola unidad en venta, lo que sugiere que son vendedores pequeños en la plataforma (*hobby sellers*). Estas publicaciones tienen bajo potencial de ingreso a *fulfillment* y no son de interés principal para la plataforma. Solo el 20% de las observaciones cuentan con más de 10 unidades a la venta al momento de la creación de la publicación, y el máximo es de 999 unidades (descartando el 1% de *outliers*).

La Figura 2.30 a la derecha analiza la relación entre las unidades a la venta al momento de la creación de la publicación y la demanda del SKU en las primeras seis semanas. No se observa una relación directa entre la cantidad de unidades disponibles y la demanda en las primeras seis semanas. Algunos SKUs con menos de 100 unidades disponibles al momento de la creación de la publicación experimentan una demanda significativamente mayor que 100 unidades en las primeras seis semanas.

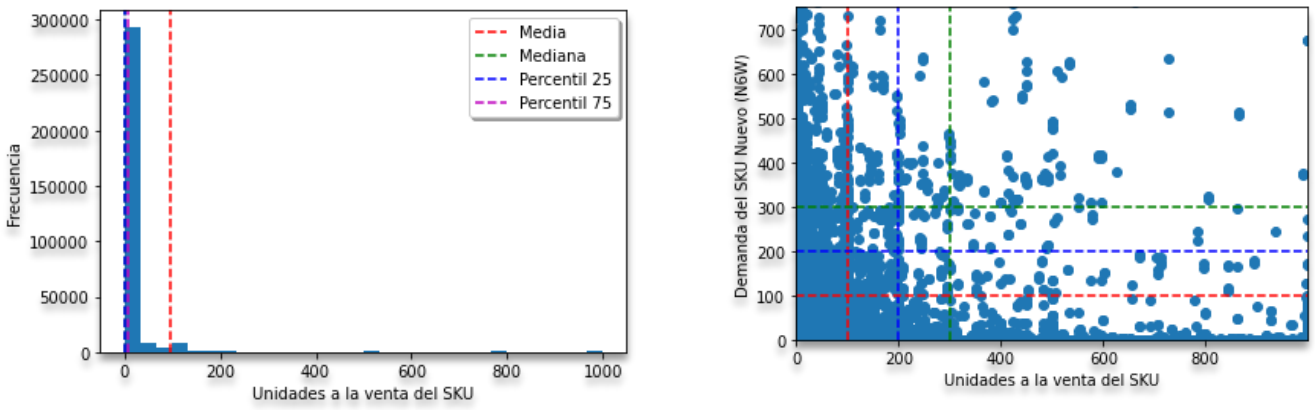


Figura 2.30 Izquierda: Observaciones según unidades a la venta al momento de la creación de la publicación. Derecha: Distribución de la demanda según cantidad de unidades disponibles a la venta al momento de la creación de la publicación

Estos hallazgos son relevantes para el modelado del problema, ya que indican que la cantidad de unidades disponibles a la venta al momento de crear la publicación no está directamente relacionada con las unidades vendidas en el plazo de las primeras seis semanas. Además, proporcionan una pista para la plataforma para distinguir entre SKUs que pertenecen a *hobby sellers* y aquellos SKUs que podrían tener potencial para ingresar a *fulfillment*.

Marca del SKU

La Figura 2.31 analiza las marcas más representativas del conjunto de datos, que concentran el 80% de las observaciones. *'Xiaomi'* es la marca más repetida, con el 30% de las observaciones, seguida de *'Apple'* con el 24% y *'Samsung'* con el 11%. Las demás marcas tienen una participación inferior al 7%.

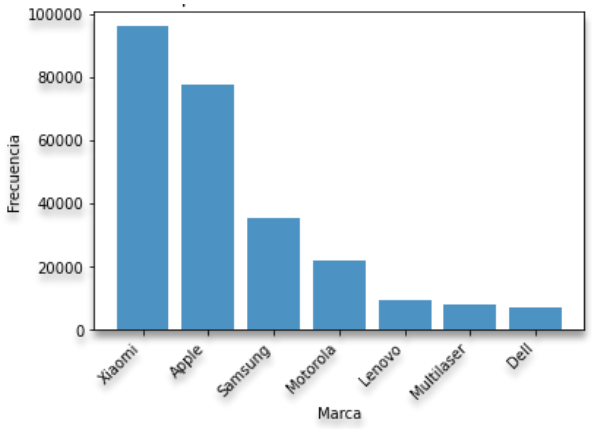


Figura 2.31 Observaciones según marca del SKU (top marcas: hasta percentil 80)

La Figura 2.32 muestra que no siempre una mayor cantidad de observaciones de una marca se traduce en una mayor demanda en las primeras seis semanas del producto. Por ejemplo, aunque el 30% de las observaciones corresponden a la marca *'Xiaomi'*, el promedio de ventas en las primeras seis

semanas de un SKU de esta marca es de 5,4. En contraste, *'Motorola'*, que solo representa el 6.7% de las observaciones, tiene un promedio de ventas de 8,1, un 50% más.

Marca	Media	Mediana	Perc'25	Perc'50	Perc'75	Perc'90	Perc'95	Perc'99
Xiaomi	5,4	0	0	0	0	6	20	124
Apple	0,6	0	0	0	0	0	2	14
Samsung	6,8	0	0	0	0	9	26	156
Motorola	8,1	0	0	0	2	11	34	173
Lenovo	3,1	0	0	0	0	3	9	67
Multilaser	6,6	0	0	0	0	7	29	165
Dell	1,2	0	0	0	0	2	3	25

Figura 2.32 Distribución de la demanda de SKU nuevo por marca

Estos hallazgos destacan la importancia de la marca en la determinación de la demanda de un producto nuevo. Por lo tanto, el *marketplace* debería considerarla como un atributo relevante al definir la política de *stock* máximo que puede ingresar un producto nuevo, ya que influye en el volumen de ventas esperado.

Línea del SKU

La Figura 2.33 a la izquierda muestra la distribución de las observaciones según la línea del SKU, centrándose en aquellas que representan el 75% de las observaciones del *dataset*. Se observa que las dos primeras líneas tienen una mayor concentración que el resto; *'iPhone'* representa el 22% del total y *'Redmi'* el 18%, mientras que las demás líneas tienen menos del 8% cada una.

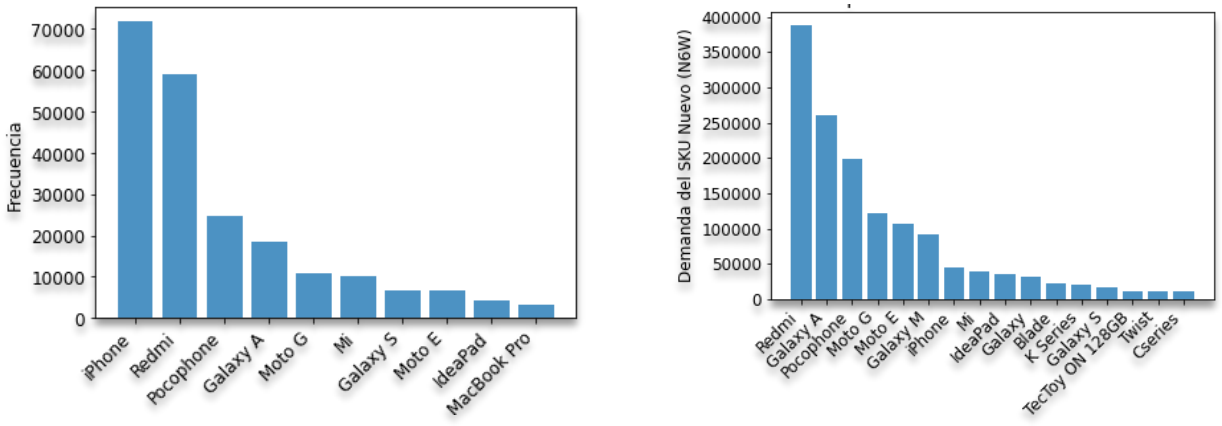


Figura 2.33 Izquierda: Distribución de observaciones según línea del SKU (75% de las observaciones). Derecha: 36 Distribución de la demanda según línea del SKU (75% de demanda)

Sin embargo, la Figura 2.33 a la derecha que analiza las líneas que concentran el mayor volumen de ventas de SKUs nuevos, presenta una imagen algo distinta. *'Redmi'* concentra 21% del total, *'Galaxy A'* un 13,8% y *'Pocophone'* un 11%. Las demás líneas concentran menos del 7% de la demanda.

La diferencia en la concentración de observaciones y demanda entre las líneas del SKU sugiere que ciertas líneas tienen un mayor potencial de ventas en las primeras seis semanas, lo que debe considerarse al determinar la cantidad de *stock* a ingresar en *fulfillment* para maximizar las ventas.

Modelo del SKU

En contraste con las variables marca o línea, la Figura 2.34 a la izquierda muestra que la distribución de los modelos de productos es muy atomizada, con las 22 principales marcas representando solo el 40% de las observaciones. Los más relevantes son *'iPhone 11'* de Apple, *'Note 10'* de Samsung y *'Poco X3 Pro'* de Xiaomi.

Analizando en la Figura 2.34 a la derecha, los modelos que concentran el 40% de la demanda de SKUs nuevos, la lista es diferente. El modelo *'9A'* lidera con el 4,4% de la demanda total, seguido por el *'Poco X3 Pro'* con un 4,3% y el *'Note 10S'* con un 3,6%.

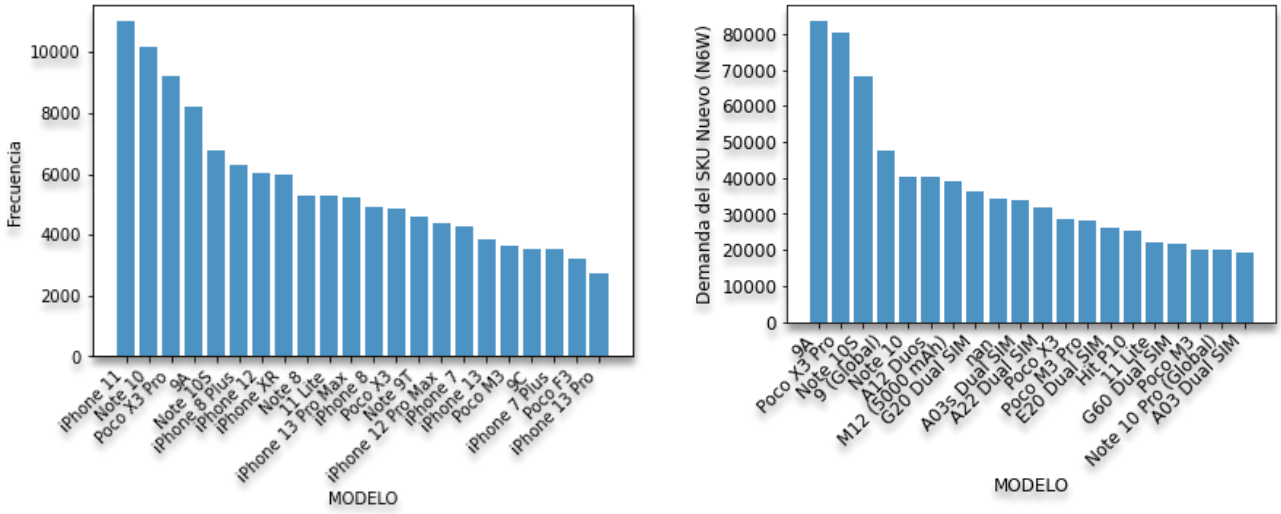


Figura 2.34 Izquierda: Distribución de observaciones según modelo del SKU (40% de las observaciones). Derecha: Distribución de la demanda según modelo del SKU (40% de demanda)

La dispersión de demanda entre diferentes modelos sugiere que la plataforma debería considerar este feature al estimar la demanda esperada y definir la política de *stock* máximo.

En la sección del Apéndice 9.4 se detallan las exploraciones entre las restantes *features* y la variable *target*.

3 Metodología

En esta sección, se describe la metodología utilizada para desarrollar y evaluar modelos predictivos en el contexto de la demanda de productos nuevos del marketplace más grande de la región. Se abordan algoritmos de aprendizaje supervisado, técnicas de selección de modelos y métodos de evaluación, incluyendo la validación cruzada y diversas métricas de rendimiento.

3.1 Algoritmos de Aprendizaje Supervisado

Hastie y Friedman[46] definen en su libro *"The Elements of Statistical Learning"* el algoritmo de aprendizaje supervisado de la siguiente manera:

> En un escenario típico, tenemos una medición de resultado, generalmente cuantitativa (como el precio de las acciones) o categórica (como ataque cardíaco/no ataque cardíaco), que deseamos predecir basándonos en un conjunto de características (como la dieta y las medidas clínicas). Tenemos un conjunto de datos de entrenamiento, en el cual observamos las medidas de resultado y características para un conjunto de objetos (como personas). Usando estos datos, construimos un modelo de predicción, o aprendiz, que nos permitirá predecir el resultado para nuevos objetos no vistos. Un buen aprendiz es aquel que predice con precisión dicho resultado. Los ejemplos anteriores describen lo que se llama el problema de aprendizaje supervisado.
>
> Se llama "supervisado" debido a la presencia de la variable de resultado para guiar el proceso de aprendizaje. En el problema de aprendizaje no supervisado, solo observamos las características y no tenemos mediciones del resultado. Nuestra tarea es describir cómo se organizan o agrupan los datos.

El objetivo de los algoritmos de aprendizaje supervisado es identificar una función matemática que establezca una relación entre las variables de entrada (*inputs*) y la variable de salida (*output*). Dicha función, conocida como modelo, se emplea para realizar predicciones sobre la variable de salida con nuevas entradas no incluidas en el conjunto de datos original. En esta tesis, se utiliza como *inputs* los metadatos de SKUs nuevos y el *output* es su demanda en las primeras seis semanas en el mercado.

Como indican los autores, existen dos tipos fundamentales de problemas en el aprendizaje supervisado: regresión y clasificación. La regresión se ocupa de variables de salida continuas con el fin de predecir un valor numérico, mientras que la clasificación maneja variables de salida discretas, buscando predecir la clase a la que pertenece una nueva entrada.

En el estudio de esta tesis, se entrenaron dos tipos de modelos: de regresión y de clasificación. Inicialmente, la variable de salida era continua (unidades demandadas de un SKU sin historial de ventas durante las primeras seis semanas), por lo que se comenzó entrenando modelos de regresión. Sin embargo, al comparar los resultados de estos modelos con los obtenidos en los modelos de clasificación, se descubrió que estos últimos ofrecían resultados significativamente mejores.

Considerando la aplicación práctica de esta información en el *marketplace* para proporcionar recomendaciones a los vendedores respecto al ingreso de inventario a *fulfillment*, se concluyó que presentar rangos de unidades inspira mayor confianza que proporcionar un valor numérico exacto. Por lo tanto, se decidió transformar la variable de salida en categórica, estableciendo rangos de demanda para el SKU nuevo en las primeras seis semanas de ventas. Esta conversión facilitó la agrupación de productos y mejoró la precisión del modelo al generar predicciones.

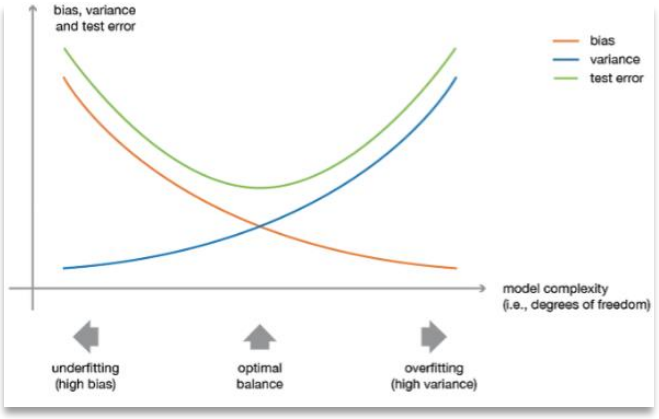
Selección de Modelo y Tradeoff entre Sesgo y Varianza

La selección del modelo para predecir la demanda de un nuevo SKU sin historial de ventas se basó en el equilibrio entre sesgo y varianza, buscando asegurar una capacidad de generalización adecuada en datos no vistos.

El sesgo es la diferencia entre la predicción promedio del modelo y el valor real de la variable objetivo. Un alto sesgo indica que el modelo no capta adecuadamente la complejidad de los datos, conocido como *underfitting*.

La varianza refiere a la variabilidad de las predicciones del modelo ante cambios en los datos de entrenamiento. Una alta varianza implica que el modelo está sobreajustado a los datos de entrenamiento y no generaliza bien a nuevos datos, lo que se conoce como *overfitting*.

El objetivo es encontrar un balance entre sesgo y varianza para minimizar el error en los datos de prueba. Un modelo óptimo debe ser complejo para capturar la estructura de los datos, pero no tanto que se ajuste a ruido o variaciones aleatorias. La Figura 3.1 ilustra el *tradeoff* entre sesgo y varianza, mostrando cómo la meta es encontrar un punto de equilibrio donde el modelo tenga un rendimiento adecuado sin caer en *underfitting* o *overfitting*.



47

Figura 3.1 Tradeoff entre sesgo, varianza y error en datos de prueba

Para abordar el desafío de estimar la demanda de un SKU nuevo, se exploraron dos técnicas de aprendizaje supervisado: árboles de decisión y modelos ensamblados (*Bagging* y *Boosting).* Los árboles de decisión son intuitivos y permiten desglosar decisiones basadas en características específicas, mientras que los modelos ensamblados combinan múltiples modelos para mejorar la precisión y reducir el riesgo de *overfitting* sin aumentar significativamente el sesgo.

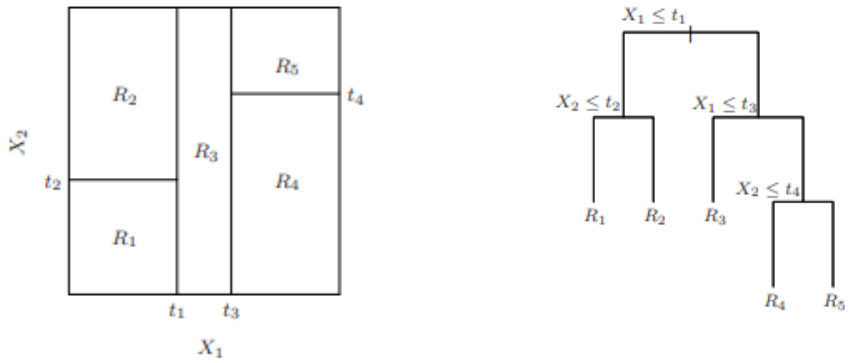
3.1.1 *Árboles de Decisión*

Los árboles de decisión, como se describe en *The Elements of Statistical Learning*[48], son herramientas esenciales en aprendizaje automático y minería de datos para predecir valores o clasificar datos en categorías basándose en un conjunto de variables independientes. A diferencia de los modelos de regresión lineal, su habilidad para procesar variables categóricas y capturar relaciones no lineales los convierte en instrumentos idóneos para predecir la demanda de SKUs nuevos sin historial de ventas.

La construcción de un árbol de decisión inicia con una pregunta sencilla, expandiéndose mediante ramificaciones que culminan en hojas representando categorías o valores medios. El algoritmo busca subdivisiones homogéneas en el conjunto de datos de entrenamiento, identificando patrones específicos basados en atributos del producto y del vendedor. El árbol resultante se aplica para realizar predicciones sobre nuevos conjuntos de datos, siguiendo el camino de preguntas hasta llegar a una hoja que ofrece la predicción deseada.

Como ejemplo, en la figura 3.2 a la izquierda se observa la partición del espacio de características mediante división binaria recursiva aplicado a datos ficticios. A la derecha, el árbol correspondiente,

donde las observaciones que cumplen con la condición en cada nodo avanzan hacia una rama determinada, y las hojas terminales representan las regiones R1, R2, ..., R5.



49

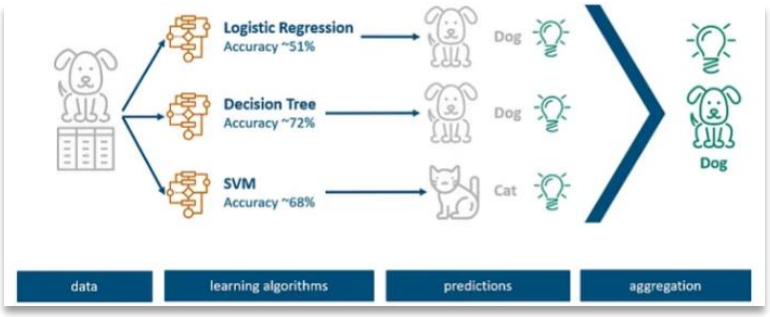
Figura 3.2. Izquierda: partición de espacio de características bidimensional mediante división binaria recursiva, aplicado a algunos datos ficticios. Derecha: árbol correspondiente a la partición del gráfico a la izquierda

Los árboles de decisión se presentan en dos variantes principales: regresión para variables continuas y clasificación para variables categóricas. Ambas variantes son útiles en este contexto. La regresión puede ser empleada para predecir la cantidad exacta de demanda de un SKU, mientras que la clasificación puede categorizar SKUs en rangos de demanda, como "baja", "media" o "alta", aportando mucha flexibilidad.

3.1.2 *Modelos Ensamblados*

Los modelos ensamblados son estrategias de aprendizaje automático que mejoran la precisión y la consistencia de las predicciones al combinar las salidas de varios modelos más simples. Este enfoque es particularmente útil cuando ningún modelo individual es consistentemente preciso. En lugar de depender de un solo modelo, los modelos ensamblados utilizan la diversidad de varios modelos para producir una predicción consolidada y más robusta.

La Figura 3.3 ilustra el concepto de un modelo ensamblado. En este ejemplo simplificado, dos modelos predicen que una observación es un perro y uno la clasifica como un gato. El enfoque ensamblado permite integrar estos diferentes puntos de vista para tomar una decisión final, equilibrando los resultados de cada modelo individual.



[50]

Figura 3.3 Ejemplo de modelo ensamblado

Esta técnica de agregación o votación se utiliza para integrar diferentes puntos de vista y tomar una decisión final más confiable, aprovechando el consenso entre modelos y reduciendo el impacto del ruido y de los sobreajustes individuales.

Para predecir la demanda de SKUs nuevos sin historial de ventas, se exploran dos técnicas de ensamblado: *Bagging* y *Boosting*. Estas técnicas son esenciales para manejar la incertidumbre y variabilidad de los productos nuevos, donde la falta de datos históricos dificulta las predicciones.

Bagging (Bootstrap Aggregation)

El *Bagging*, es una técnica para reducir la varianza y mitigar el *overfitting* de los modelos predictivos como árboles de decisión y regresión. Consiste en entrenar múltiples modelos en paralelo utilizando subconjuntos aleatorios del conjunto de datos de entrenamiento, conocidos como "*bags*".

Según Dieckmann[44], la fuerza del *Bagging* reside en que, al entrenar cada modelo con diferentes segmentos de datos, sus predicciones individuales se combinan para formular una predicción final más robusta y precisa. En problemas de regresión, se promedian las predicciones de los modelos, mientras que en clasificación se opta por un sistema de votación para establecer la clase predominante.

Dentro de este enfoque, el *Random Forest* es un modelo basado en B*agging* que, además, aleatoriza las *features* durante el entrenamiento de cada árbol. Esto significa que cada árbol se entrena no solo con un subconjunto aleatorio de los datos, sino también con un subconjunto aleatorio de las *features*, lo que reduce la varianza sin aumentar el sesgo.

Este enfoque es especialmente útil para predecir la demanda de SKUs nuevos sin historial de ventas, manejando bien la complejidad y la variabilidad de los datos limitados.

Boosting

Boosting es una técnica de ensamblado de modelos mejora la precisión de los modelos al reducir el sesgo y evitar el *underfitting*. A diferencia del *Bagging*, que reduce la varianza entrenando modelos independientes, *Boosting* construye modelos secuenciales que aprenden de los errores de los anteriores, enfocándose en reducir el sesgo.

Según Dieckmann[51], el objetivo principal de *Boosting* es aumentar la precisión del modelo combinando múltiples modelos débiles en un modelo fuerte y consistente. El proceso de *Boosting* implica entrenar secuencialmente varios modelos débiles, cada uno enfocado en corregir los errores de su predecesor, y luego combinarlos para formar un modelo final más preciso. El proceso se resume en tres pasos:

1. Entrenamiento inicial: se entrena un modelo débil con un subconjunto de los datos.
2. Asignación de pesos: se entrena un nuevo modelo débil en un subconjunto aleatorio, asignando mayor peso a las observaciones que fueron mal clasificadas por el modelo anterior.
3. Modelo ensamblado: el siguiente modelo se enfoca en mejorar el rendimiento en las observaciones difíciles y se combina con los modelos anteriores en un nuevo modelo ensamblado.

Este ciclo se repite hasta lograr un nivel de precisión aceptable.

En la práctica, técnicas como *Gradient Boosting* y *XGBoost* son algoritmos populares para ensamblar árboles de decisión.

En esta tesis, se utilizó *XGBoost* para estimar la demanda de SKUs nuevos sin historial de ventas debido a su capacidad para manejar datos de alta dimensionalidad y su eficiencia computacional. *XGBoost* permite aprender de SKUs similares y hacer predicciones precisas para productos nuevos en el *marketplace*.

3.2 Técnicas de Evaluación del Modelo

3.2.1 *Métricas de Performance*

Mean Absolute Percentage Error (MAPE)

En la presente tesis, se utilizaron dos métricas principalmente para evaluar el rendimiento de los modelos entrenados: MAPE y *F-1 score*.

El MAPE *(Mean Absolute Percentage Error)* se utilizó para medir los resultados del modelo de regresión, que estimaba la demanda de productos nuevos en las primeras seis semanas como un valor discreto. El MAPE mide la precisión de modelos de regresión en términos porcentuales, interpretándose como el porcentaje promedio de desviación entre las predicciones del modelo y los valores reales. Un valor menor de MAPE indica un mejor desempeño del modelo.

Para calcular el MAPE, se mide el error absoluto de cada observación, es decir, la diferencia entre la predicción del modelo ($\hat{y}_i$) y el valor real de dicha observación (y_i). Esta diferencia se convierte en porcentaje dividiéndola por el valor real y multiplicándola por 100. Luego, se promedian todos los errores absolutos porcentuales. Matemáticamente, se expresa así:

$$MAPE = \frac{1}{n} * \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} * 100 \right|$$

Entre las ventajas de utilizar el MAPE como métrica de evaluación se destaca la simplicidad de interpretación al expresarse en porcentaje, lo cual facilita la comparación entre modelos. Además, es robusto ante valores atípicos, ya que se calcula en términos porcentuales y no en unidades absolutas.

Sin embargo, el MAPE también presenta ciertas desventajas. No considera la dirección del error, es decir, no diferencia entre sobreestimación y subestimación. Además, no es adecuado para problemas con valores iguales a cero, ya que el denominador de la ecuación resulta en un valor infinito[52].

En esta tesis, el MAPE fue elegido por su capacidad para interpretar el rendimiento del modelo en términos relativos y facilitar la comparación entre modelos. Sin embargo, debido a que el 80% de las observaciones tenían valores iguales a cero (ver Figura 3.4), los modelos de *Random Forest* y *XGBoost Regressor* mostraron resultados muy altos de MAPE.

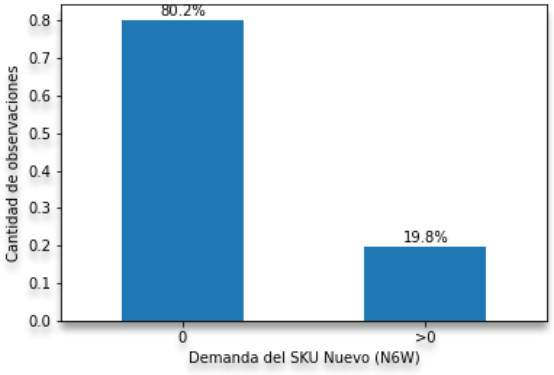


Figura 3.4 Distribución de observaciones según la variable target

Accuracy

El bajo rendimiento del modelo de regresión inicial llevó a transformar la variable objetivo en una categórica y a evaluar métricas de clasificación. Se consideró la métrica de *Accuracy* por su simplicidad y facilidad de interpretación, definida como:

$$Accuracy = \frac{Nuber\ of\ correct\ predictions}{Total\ number\ of\ predictions} = \frac{TP + TN}{(TN + FP + FN + TP)}$$

Para calcular *Accuracy*, es fundamental comprender la matriz de confusión (Figura 3.5), que clasifica las predicciones del modelo en cuatro categorías: Verdaderos Positivos (TP), Falsos Positivos (FP), Verdaderos Negativos (TN) y Falsos Negativos (FN)

	$y = 1$	$y = 0$
$\hat{y} = 1$	TP/N_+=TPR=sensitivity=recall	FP/N_-=FPR=type I
$\hat{y} = 0$	FN/N_+=FNR=miss rate=type II	TN/N_-=TNR=specifity

53

Figura 3.5 Estimación de p(ŷ|y) a partir de una matriz de confusión

Métricas derivadas de la matriz de confusión:

1. *True Positive Rate (TPR):* también conocido como *recall* o sensibilidad, indica la proporción de positivos reales correctamente clasificados.

$$TPR = \frac{TP}{N_+} = P(\hat{y} = 1 | \, y = 1)$$

2. *False Positive Rate (FPR):* también llamado error de tipo I, representa la proporción de negativos reales que se clasificaron incorrectamente como positivos.

$$FPR = \frac{FP}{N_-} = P(\hat{y} = 1 | \, y = 0)$$

3. *False Negative Rate (FNR)*: también conocido como error de tipo II, indica la proporción de positivos reales que se clasificaron incorrectamente como negativos.

$$FNR = \frac{FN}{N_+} = P(\hat{y} = 0 | \, y = 1)$$

4. *True Negative Rate (TNR):* también llamado especificidad, representa la proporción de negativos reales que se clasificaron correctamente como negativos.

$$True\ Negative\ Rate = \frac{TN}{N_-} = P(\hat{y} = 0 | \, y = 0)$$

Sin embargo, debido al desbalance de clases observado en los datos de esta investigación, no se utilizó la métrica de *Accuracy* para evaluar el rendimiento del modelo. La alta concentración de observaciones en la clase "0-10" podría conducir a conclusiones engañosas, ya que un modelo que predice la clase mayoritaria podría mostrar una precisión artificialmente alta. Además, *Accuracy* no considera la ponderación de las clases según su importancia relativa, lo cual es crucial en este contexto donde se busca un enfoque diferenciado para la clase que se espera tenga ventas superiores al límite actual de 20 unidades por producto.

Para una evaluación más precisa del modelo, se optó por métricas como el *F1-score* y matrices de probabilidad por clase, que abordan el desbalance de clases y la importancia relativa de cada clase en los objetivos de la investigación.

F1-score

El *F1-score* es una métrica crítica para evaluar el rendimiento de modelos de clasificación con un umbral (τ) fijo. Combina *precision* y *recall* en una sola medida para equilibrar la capacidad del modelo de identificar correctamente observaciones positivas y negativas, evitando sesgos.

El umbral de clasificación (τ) determina cómo se asignan las predicciones a las clases positiva y negativa. Ajustar este umbral permite equilibrar la sensibilidad y la especificidad del modelo.

El *F1-score* oscila entre 0 y 1, donde 1 representa un rendimiento perfecto y 0 indica un rendimiento muy deficiente. Su fórmula de cálculo es:

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall}$$

En el contexto de modelos de clasificación multiclase, como el abordado en esta tesis, se emplea la estrategia *"one vs. all",* que consiste en entrenar múltiples clasificadores binarios independientes, cada uno diseñado para distinguir una clase específica del resto. Según la bibliografía de *scikit-learn*[54], este enfoque no solo es eficiente computacionalmente, sino que también facilita la interpretación y evaluación del modelo para cada clase individualmente.

Para calcular *F1-score* es relevante evaluar si hay desequilibrio entre las clases. Como se observa en la Figura 3.6, en el *dataset* analizado la clase "0-10" concentra aproximadamente el 94% de las observaciones, mientras que las clases "11-20" y ">21" concentran el 2.1% y el 3.9%, respectivamente.

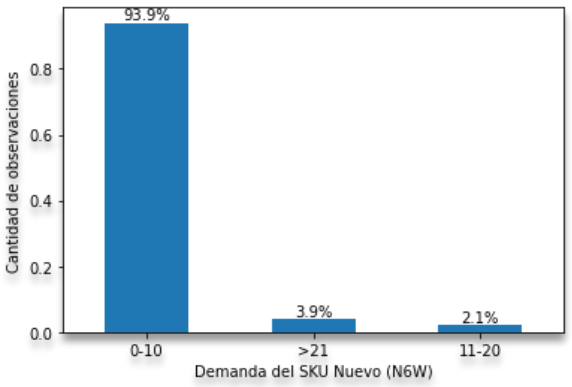


Figura 3.6 Distribución de la variable objetivo

Para abordar este desequilibrio, se utilizó una variante del *F1-score* conocida como *F1-score* ponderado, que asigna pesos a cada clase según su proporción en los datos. Esta medida ajustada permite una evaluación más precisa del rendimiento del modelo, considerando adecuadamente la distribución desigual de las clases en los datos de entrada. Se calcula como:

$$Weighted\ F1 = \sum_{i=1}^{N} w_i * F1_i$$

Donde:

$$w_i = \frac{number\ of\ samples\ in\ class\ i}{total\ number\ of\ samples}$$

Matriz de Probabilidad

La matriz de probabilidad se utilizó para mejorar la toma de decisiones basadas en las predicciones del modelo desarrollado, proporcionando una herramienta efectiva para adaptar las decisiones según la confianza en cada predicción.

Una matriz de probabilidad es una tabla que muestra las clases en las filas y los rangos de probabilidad en las columnas. Cada celda indica la proporción de instancias de una clase específica que caen en un rango de probabilidad determinado. Esto permite analizar cómo el modelo hace predicciones en diferentes niveles de confianza.

Clase	Prob 0-10	Prob 10-20	Prob 20-30	...	Prob 90-100
0-10	15%	12%	10%	...	5%
11-20	5%	8%	12%	...	20%
>21	2%	4%	8%	...	30%

Figura 3.7 Ejemplo de matriz de deciles

En la matriz presentada en la Figura 3.7, cada fila representa una clase de ventas (0-10 unidades, 11-20 unidades, y más de 21 unidades), mientras que cada columna representa un decil de confianza en

las predicciones del modelo. Por ejemplo, se observa que el 30% de las predicciones para la clase ">21 unidades" caen en el decil 10, indicando un alto nivel de confianza en estas predicciones.

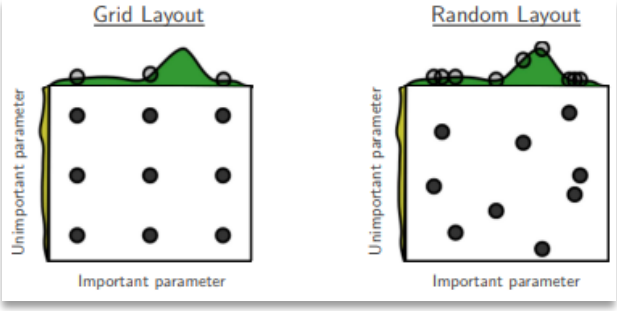
La matriz de probabilidad fue crucial para tomar decisiones diferenciadas por clase, dependiendo del nivel de confianza o riesgo deseado. Por ejemplo, para la clase que representa ventas superiores al límite permitido (>21 unidades), se estableció un umbral más conservador en los deciles más altos. Esto permite identificar con mayor certeza los casos que requieren una revisión más exhaustiva.

3.2.2 *Optimización de Hiperparámetros*

Para encontrar los mejores hiperparámetros para cada modelo, se utilizó la función '*RandomizedSearchCV*' de *scikit-learn.* Esta función realiza una búsqueda aleatoria dentro de un espacio de hiperparámetros especificado, seleccionando combinaciones aleatorias y evaluando el rendimiento del modelo para cada una, siendo más eficiente que una búsqueda exhaustiva.

El proceso comienza seleccionando combinaciones de hiperparámetros al azar, entrenando y evaluando el modelo con cada combinación. Se usa validación cruzada dividiendo los datos en 5 pliegues y evaluando el rendimiento del modelo con uno de estos pliegues como conjunto de validación. La mejor combinación de hiperparámetros se selecciona según el mejor rendimiento promedio.

A diferencia de '*GridSearchCV*', '*RandomizedSearchCV*' explora un número fijo de configuraciones de parámetros, determinado por *n_iter*, que en este caso fue 10. Esto permite una cobertura más eficiente del espacio de hiperparámetros, como se ilustra en la Figura 3.8.



55

Figura 3.8 Cobertura del espacio de hiperparámetros: cuadrícula de puntos vs. puntos aleatorios

En esta tesis, se exploraron diversos hiperparámetros para los modelos *Random Forest* y *XGBoost*, tanto en regresión como en clasificación.

Random Forest (Regressor / Classifier):

- *max_depth:* profundidad máxima de cada árbol. A mayor profundidad, captura relaciones más complejas, pero puede sobreajustar. Valores explorados: [5, 7, 10, 20].
- *n_estimators:* número de árboles en el modelo. Más árboles aumentan la robustez, pero también el tiempo de entrenamiento. Valores explorados: [5, 10, 25, 50].
- *min_samples_split:* número mínimo de observaciones para dividir un nodo interno. Evita divisiones innecesarias y reduce el sobreajuste. Valores explorados: [2, 5, 10].
- *min_samples_leaf:* número mínimo de observaciones en una hoja (nodo final). Similar a *min_samples_split*, previene divisiones excesivas. Valores explorados: [1, 2, 4].
- *max_features:* número máximo de *features* consideradas para cada división del árbol. Valores explorados: ['auto', 'sqrt', 'log2', None]. 'Auto' y 'sqrt' corresponden a la raíz cuadrada del total, 'log2' al logaritmo en base 2 del total y 'None' considera todas.

XGBoost (Regressor / Classifier):

- *max_depth:* profundidad máxima de cada árbol. Valores explorados: [5, 7, 10, 20].
- *n_estimators*: número de árboles en el modelo. Valores explorados: [5, 10, 25, 50].
- *learning_rate:* tasa de aprendizaje del algoritmo. Un valor más bajo permite un aprendizaje más gradual para evitar el sobreajuste, pero puede requerir más iteraciones para alcanzar precisión. Valores explorados: [0.05, 0.1, 0.2, 0.5, 1].

3.2.3 *Validación Cruzada*

En la construcción de un modelo de aprendizaje automático, es crucial estimar su rendimiento en datos desconocidos, es decir, observaciones que el modelo no ha visto previamente. Si se entrena un modelo con un conjunto de datos de entrenamiento y luego se utilizan los mismos datos para evaluar su rendimiento, se incurre en una evaluación sesgada.

La validación cruzada es esencial para evaluar la capacidad de generalización de un modelo y evitar tanto el subajuste (modelo demasiado simple) como el sobreajuste (modelo demasiado complejo). Esta técnica ayuda a equilibrar el sesgo y la varianza, asegurando un buen desempeño en datos nuevos. Según Hastie, Tibshirani y Friedman en *The Elements of Statistical Learning*, “probablemente el método más simple y ampliamente utilizado para estimar el error de predicción es la validación cruzada”[56].

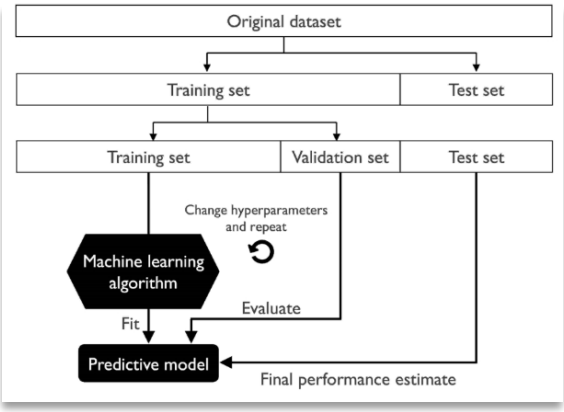
Dos técnicas comunes de validación cruzada son el método *Holdout* y *K-fold.*

Método Holdout de Validación Cruzada

El método de *Holdout* divide el conjunto de datos en dos partes: entrenamiento y prueba. l conjunto de entrenamiento se utiliza para entrenar el modelo, mientras que el de prueba evalúa su rendimiento en datos desconocidos.

Para probar diferentes configuraciones de hiperparámetros, se recomienda una variante que divide los datos en tres partes: entrenamiento, validación y prueba (ver Figura 3.9). El conjunto de entrenamiento se usa para ajustar distintos modelos, el de validación evalúa su rendimiento para seleccionar el mejor modelo, y el de prueba mide el rendimiento del modelo en datos no vistos previamente, proporcionando una estimación menos sesgada de la capacidad de generalización del modelo.

Es importante considerar que el método *Holdout* es sensible a cómo se dividen los datos en los conjuntos de entrenamiento y validación.



57

Figura 3.9 Representación de metodología holdout

Método K-fold de Validación Cruzada

En la validación cruzada *K-fold*, los datos se dividen aleatoriamente en K pliegues. Se utilizan K-1 pliegues para entrenar el modelo y el pliegue restante para evaluarlo. Este proceso se repite K veces, obteniendo K modelos y estimaciones de rendimiento independientes. Se calcula el rendimiento promedio, proporcionando una estimación menos sensible a la subdivisión de los datos en comparación con el método *Holdout*.

Esta técnica se usa generalmente para ajustar hiperparámetros. Una vez optimizados, el modelo se entrena con todos los datos de entrenamiento y se evalúa con un conjunto de prueba independiente, produciendo un modelo más preciso y robusto.

En esta tesis, se implementaron tanto el método *Holdout* como el método *K-fold* para entrenar y evaluar el modelo de estimación de demanda de SKUs nuevos sin historial de ventas. Primero, se dividieron los datos utilizando el método *Holdout*, asignando el 80% al conjunto de entrenamiento y el 20% al conjunto de prueba, como se muestra en la Figura 3.10. Para ello, se utilizó *'GroupShuffleSplit'* de *scikit-learn*, agrupando los elementos por vendedor, garantizando que los datos de un mismo vendedor no se dividan entre ambos conjuntos. Esta estrategia previene sesgos y asegura una mejor generalización del modelo en datos desconocidos.

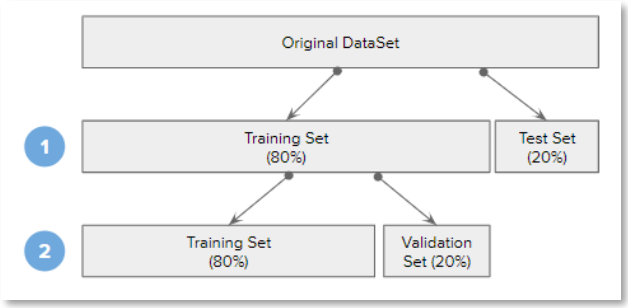


Figura 3.10 Criterios de proporciones utilizados para la validación del modelo

Aunque se consideró mantener productos idénticos en conjuntos separados, esta opción fue descartada ya que cada producto en el *dataset* es único en su configuración, y no existen dos productos exactamente iguales. Cada producto representa una combinación única de características como tipo de producto, marca, modelo, línea, color, vendedor y fecha de publicación. Dadas estas particularidades, mantenerlos en conjuntos separados no tendría sentido y no aportaría ningún beneficio adicional al proceso de entrenamiento y evaluación del modelo. Por lo tanto, se optó por no implementar esta separación para los productos.

Para optimizar los hiperparámetros y obtener una estimación robusta, se aplicó la validación cruzada *K-fold* con 5 pliegues. Cada pliegue actuó como conjunto de prueba una vez, repitiéndose el proceso 5 veces y obteniendo una estimación promedio del rendimiento del modelo.

Estas metodologías permitieron una evaluación justa y rigurosa del modelo, asegurando su capacidad para generalizar en datos desconocidos y proporcionando una base sólida para la selección de hiperparámetros óptimos.

3.3 Técnicas de *Feature Engineering*

Para el preprocesamiento de datos, se utilizó la biblioteca *scikit-learn,* aplicando diversas técnicas de *feature engineering* tanto a las variables numéricas como a las categóricas:

- **Relleno de valores faltantes:** para las variables numéricas, se rellenaron los valores faltantes (NaN) con el valor constante cero. Para las variables categóricas, se rellenaron los valores faltantes con la etiqueta "*missing*".
- **Normalización de variables numéricas:** se normalizaron ciertas variables numéricas que tenían diferentes unidades de medida (como *'ancho', 'altura', 'profundidad', 'volumen'* y *'peso'*) para asegurar una correcta comparación entre las observaciones.
- **Codificación de variables categóricas*:*** se aplicó la técnica de codificación *One Hot Encoding* a las variables categóricas para transformarlas en representaciones numéricas binarias.
- **Simplificación de categorías:** se simplificó la cantidad de opciones que podían tomar algunas variables categóricas relevantes para ayudar al modelo a comprender las relaciones entre las observaciones. Por ejemplo, se unificaron todas las variantes de un mismo color bajo una única etiqueta y se agruparon las opciones restantes bajo la categoría "otros".
- **Normalización y limpieza de variables categóricas específicas:** Se realizó una normalización y limpieza de las variables *'marca', 'modelo'* y *'línea'* para asegurar una representación coherente y limpia de estas características.

Estas técnicas de *feature engineering* fueron fundamentales para garantizar que los datos estuvieran en un formato adecuado para entrenar el modelo de *machine learning.*

3.4 *Software*

Este proyecto ha sido desarrollado íntegramente utilizando el lenguaje de programación *Python*, con la utilización del entorno *Jupyter Notebook,* que ofrece una plataforma interactiva para la investigación y desarrollo de algoritmos de aprendizaje automático.

En términos de algoritmos implementados, se llevaron a cabo entrenamientos de modelos de regresión y clasificación, en particular, se utilizaron las implementaciones de *Random Forest* y *XGBoost* disponibles en la biblioteca *scikit-learn* para *Python*. El algoritmo *Random Forest* es reconocido por su capacidad para manejar conjuntos de datos ruidosos y su resistencia al sobreajuste, entre otras ventajas.

Por otro lado, *XGBoost* se destaca por su manejo de datos faltantes, capacidad para lidiar con relaciones no lineales y su eficiencia computacional en comparación con otros métodos.

Además de los algoritmos mencionados, se utilizaron bibliotecas de código abierto en *Python* para tareas relacionadas con el aprendizaje automático y procesamiento de datos, entre ellas:

- ***Matplotlib***: para la creación de gráficos y visualización de datos.
- ***Numpy***: para operaciones numéricas y manipulación de matrices.
- ***Pandas***: para la manipulación y análisis de datos tabulares.
- ***Seaborn***: para la mejora de la visualización de datos y la creación de gráficos estadísticos.
- ***Scikit-learn***: para el uso de algoritmos de aprendizaje automático y la modelización.
- ***Scikit-plot***: para la visualización de curvas ROC y matrices de confusión.

3.5 Metodología de Desarrollo

El desarrollo del modelo de la presente tesis se llevó a cabo utilizando una metodología ágil con enfoque iterativo e incremental. La metodología ágil es un enfoque de gestión de proyectos que se caracteriza por la flexibilidad, la adaptabilidad y la entrega rápida de resultados. Este enfoque permite avanzar en el desarrollo del modelo en ciclos repetitivos, obteniendo resultados tangibles de manera rápida. Además, brinda la oportunidad de mejorar continuamente el modelo a través de la retroalimentación y la evaluación constante en cada iteración.

Un ejemplo concreto de esta metodología fue el inicio del desarrollo con un enfoque de regresión. Tras completar la creación de la base de datos, preprocesamiento, entrenamiento y evaluación, se detectó rápidamente la baja performance del modelo, permitiendo reevaluar la estrategia y cambiar a un enfoque de clasificación más adecuado.

El enfoque incremental permitió construir el modelo de manera gradual, agregando funcionalidad y mejorándolo continuamente. Esto posibilitó la obtención de resultados parciales de forma temprana y ayudó a identificar y abordar rápidamente los desafíos surgidos durante el proceso. Por ejemplo, al estimar inicialmente el modelo de clasificación con 4 clases, se observó un desempeño insatisfactorio. Gracias a este enfoque, el modelo se adaptó rápidamente a 3 clases, lo que resultó en una mejor performance y calidad de los resultados.

4 Resultados

En esta sección, se analizan los resultados de los modelos de regresión y clasificación para predecir la demanda de productos sin historial de ventas. El modelo seleccionado fue un clasificador XGBoost de tres clases, que demostró la mejor capacidad para identificar la demanda esperada con un F1-score de 0.921.

4.1 Experimentación de Modelos de Regresión

A continuación, en las Figuras 4.1 y 4.2, se presentan los resultados obtenidos de dos modelos de regresión: *Random Forest Regressor* y *XGBoost Regressor.*

	Modelos de Regresión				
	Descripción del modelo	Algoritmo	N° *features*	*MAPE (train)*	*MAPE (test)*
1	Regresión optimizada	*Random Forest Regressor*	31	1.02×10^{16}	1.64×10^{16}
2	Regresión optimizada	*XGBoost regressor*	31	1.82×10^{15}	1.06×10^{16}

Figura 4.1 Resultados de los modelos de regresión

	Modelos de Regresión			
	Descripción del modelo	Algoritmo	Hiperparámetros a optimizar	Hiperparámetros optimizados
1	Regresión optimizada	*Random Forest Regressor*	*max_depth*: [5, 7, 10, 20] *n_estimators:* [5, 10, 25, 50] *min_samples_split:* [2, 5, 10] *min_samples_leaf:* [1, 2, 4] *max_features*: ['auto','sqrt','log2', None]	*max_depth* =10 *n_estimators*=5 *min_samples_split*=5 *min_samples_leaf*=1 *max_features*='log2'
2	Regresión optimizada	*XGBoost Regressor*	*max_depth*: [5, 7, 10, 20] n_estimators: [5, 10, 25, 50] *learning_rate: [*0.05, 0.1, 0.2, 0.5, 1]	*max_depth*=20 *n_estimators*=5 *learning_rate*=0.05

Figura 4.2 Hiperparámetros explorados y optimizados

4.1.1 *Modelo 1: Random Forest Regressor*

Rendimiento del modelo

Los resultados del modelo *Random Forest Regressor* son desalentadores, con valores de MAPE extremadamente altos: 1.02×10^{16} en el conjunto de entrenamiento y 1.64×10^{16} en el conjunto de prueba. Esto indica una discrepancia significativa entre las predicciones del modelo y los datos reales, mostrando una falta de precisión y capacidad del modelo para capturar los patrones de los datos.

Optimización de parámetros

El análisis de los parámetros optimizados del modelo revela algunas elecciones interesantes:

- *max_depth* = 10, buscando equilibrio para evitar sobreajuste.
- *n_estimators* = 5, prioriza la eficiencia, aunque con el riesgo de afectar el rendimiento.
- *min_samples_split* = 5, un valor intermedio que busca evitar el sobreajuste.
- *min_samples_leaf* = 1, sugiere un enfoque hacia una estructura más flexible del modelo.
- *max_features* = *'log2'*, introduciendo aleatoriedad para mejorar la generalización.

Análisis de resultados

Pese a los esfuerzos por optimizar los parámetros, la Figura 4.3 muestra que las predicciones del modelo no se alinean con los valores reales, corroborado por los altos valores de MAPE en ambos conjuntos. La línea roja, que simboliza una predicción perfecta (y=x), no muestra ningún tipo de alineamiento con los puntos de datos.

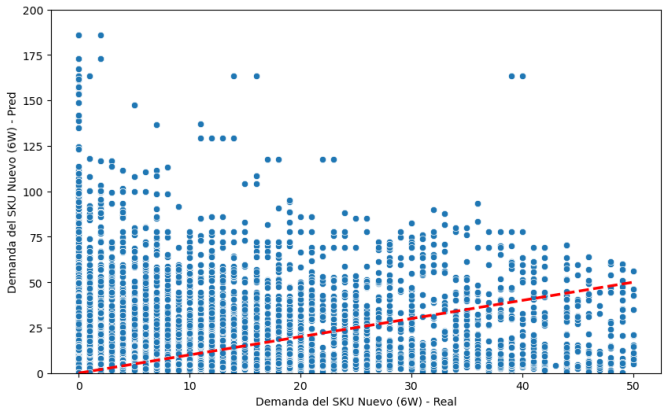


Figura 4.3 Scatterplot de la relación entre las predicciones del modelo y los valores reales

En la sección del Apéndice 9.5.1 se analiza el *feature importance* del modelo.

Conclusiones

En vista de los resultados obtenidos, se concluye que el modelo *Random Forest Regressor* no es adecuado para estimar la demanda de productos sin historial de ventas. Por lo tanto, se explorará un modelo alternativo, *XGBoost Regressor,* para abordar este problema de manera más efectiva.

4.1.2 *Modelo 2: XGBoost Regressor*

Rendimiento del modelo

El modelo *XGBoost Regressor* presenta un MAPE de 1.82×10^{15} en el conjunto de entrenamiento y 1.06×10^{16} en el conjunto de prueba. Si bien estos valores son menores que los del *Random Forest Regressor*, aún evidencian un error alto y dificultades para realizar predicciones precisas.

Optimización de parámetros

Se utilizó *'RandomizedSearchCV'* para optimizar los parámetros, resultando en:

- *max_depth* = 20, el valor máximo entre las opciones.
- *n_estimators* = 5, el valor más bajo entre las opciones.
- *learning_rate* = 0.05, el valor más bajo entre las opciones.

Estas elecciones reflejan un esfuerzo por mejorar el rendimiento del modelo minimizando el riesgo de sobreajuste. Sin embargo, el modelo aún enfrenta desafíos significativos para predecir la demanda de productos sin historial de ventas.

Análisis de resultados

La Figura 4.4, que muestra la relación entre las predicciones del modelo y los valores reales, no revela una correlación clara. Esto confirma que el modelo *XGBoost Regressor* no es adecuado para estimar la demanda de productos sin historial de ventas a partir de los datos de entrenamiento.

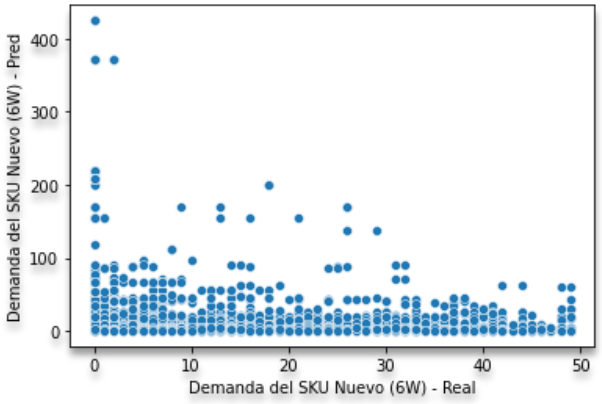


Figura 4.4 Scatterplot de la relación entre las predicciones del modelo y los valores reales

En la sección del Apéndice 9.5.2 se explora el *feature importance* del modelo.

Conclusiones

Los altos valores de MAPE revelan la incapacidad de ambos modelos para capturar con precisión las relaciones entre las características y la variable objetivo, indicando que ninguno es adecuado para abordar eficazmente el problema planteado. En vista de estos resultados, se decidió cambiar el enfoque y adaptar la variable objetivo a una de clasificación para abordar el problema de manera más efectiva.

4.2 Experimentación de Modelos de Clasificación

La conversión del problema a una tarea de clasificación es adecuada para evaluar la eficacia de la regla de negocio que permite a los vendedores ingresar entre 0 y 20 unidades de un producto nuevo en *fulfillment*. El objetivo es determinar si esta regla es adecuada o si es necesario establecer rangos más precisos para optimizar el inventario y maximizar las ventas.

Los resultados de los modelos de clasificación ajustados, *Random Forest Classifier* y *XGBoost Classifier,* en sus distintas variantes de 4 y 3 clases, se presentan en las Figuras 4.5 y 4.6.

	Modelos de Clasificación				
	Descripción del modelo	Algoritmo	*N° features*	*F1 weighted (train)*	*F1 weighted (test)*
3	Clasificación optimizada (4 clases)	*Random Forest Classifier*	31	0.938	0.909
4	Clasificación optimizada (4 clases)	*XGBoost Classifier*	31	0.952	0.914
5	Clasificación optimizada (3 clases)	*Random Forest Classifier*	31	0.961	0.919
6	Clasificación optimizada (3 clases)	*XGBoost Classifier*	31	0.998	0.919
7	Clasificación optimizada control *overfitting* (3 clases)	*XGBoost Classifier*	31	0.957	0.921

Figura 4.5 Resultados de los modelos de clasificación

	Modelos de Clasificación			
	Descripción del modelo	Algoritmo	Hiperparámetros a optimizar	Hiperparámetros optimizados
3	Clasificación optimizada (4 clases)	*Random Forest Classifier*	*max_depth:* [5, 7, 10, 20] *min_samples_leaf:* [1, 2, 4] *min_samples_split:* [2, 5, 10] *n_estimators:* [5, 10, 25, 50] *max_features:* ['auto', 'sqrt', 'log2', None]	*max_depth*=10 *min_samples_leaf*=2 *min_samples_split*=2 *n_estimators*=10 *max_features*=None
4	Clasificación optimizada (4 clases)	*XGBoost Classifier*	*max_depth:* [5, 7, 10, 20] *n_estimators:* [5, 10, 25, 50] *learning_rate:* [0.05, 0.1, 0.2, 0.5, 1]	*max_depth*=7 *n_estimators*=10 *learning_rate*=1
5	Clasificación optimizada (3 clases)	*Random Forest Classifier*	*max_depth:* [5, 7, 10, 20] *min_samples_leaf:* [1, 2, 4] *min_samples_split:* [2, 5, 10] *n_estimators*: [5, 10, 25, 50] *max_features*: ['auto', 'sqrt', 'log2', None]	*max_depth*=20 *min_samples_leaf*=4 *min_samples_split*=2 *n_estimators*=25 *max_features*=sqrt
6	Clasificación optimizada (3 clases)	*XGBoost Classifier*	*max_depth:* [5, 7, 10, 20] *n_estimators:* [5, 10, 25, 50] *learning_rate*: [0.05, 0.1, 0.2, 0.5, 1]	*max_depth*=20 *n_estimators*=25 *learning_rate*=0.5
7	Clasificación optimizada control *overfitting* (3 clases)	XGBoost Classifier	*max_depth*: [7, 9, 11, 13] *n_estimators*: [5, 7, 10] *learning_rate:* [0.4, 0.45, 0.5]	*max_depth*=9 *n_estimators*=10 *learning_rate*=0.5

Figura 4.6 Hiperparámetros explorados y optimizados

4.2.1 *Modelo 3: Random Forest Classifier con Cuatro Clases*

El modelo 3, un *Random Forest Classifier* con cuatro clases, se diseñó para evaluar y optimizar la regla de negocio actual para el ingreso de productos nuevos en *fulfillment.* Las clases se definieron de la siguiente manera:

- **Clase 0-10:** SKUs nuevos que suelen vender menos de la mitad del límite permitido por la regla actual. Se recomienda limitar su ingreso a 10 unidades para optimizar el espacio. Representa el 93.9% de las observaciones.
- **Clase 11-20:** SKUs nuevos que venden hasta 20 unidades en las primeras seis semanas, manteniendo la experiencia actual. Representa el 2.1% de las observaciones.
- **Clase 21-50:** SKUs nuevos que pueden beneficiarse al permitirles ingresar más stock, ya que venden hasta 2.5 veces más que el límite actual. Representa el 2.0% de las observaciones.
- **Clase ≥ 51:** SKUs más perjudicados por la regla actual, con alto costo de oportunidad en ventas potenciales. Representa el 2% restante de las observaciones.

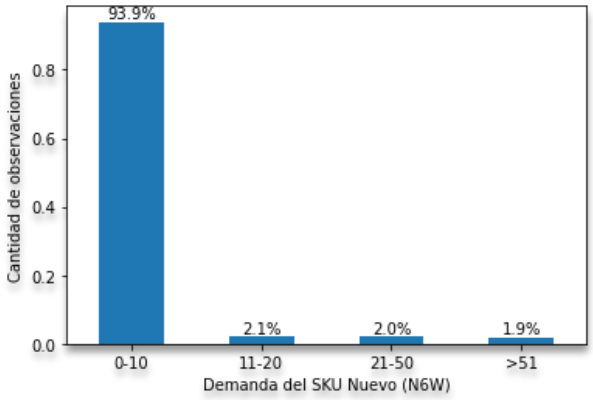


Figura 4.7 Distribución de la variable target

Rendimiento del modelo

El modelo logró un *F1-score* de 0.938 en el conjunto de entrenamiento y 0.909 en el conjunto de prueba, indicando un buen desempeño general y una capacidad adecuada para clasificar las observaciones en las cuatro clases.

En la literatura, no existe un umbral universal para determinar cuándo una diferencia entre *F1-score* en el conjunto de entrenamiento y el conjunto de prueba indica sobreajuste. La sensibilidad al sobreajuste depende de varios factores, como la complejidad del modelo y el tamaño del conjunto de datos. Sin embargo, algunas pautas generales indican que si la diferencia entre el *F1-score* de entrenamiento y test es menor al 5%, la diferencia es pequeña y probablemente no hay sobreajuste. Si la diferencia ronda entre el 5% y 10%, se considera moderada y podría indicar un ligero sobreajuste, en cuyo caso se podría ajustar el modelo o recopilar más datos. Si la diferencia es superior al 10%, probablemente estemos en presencia de sobreajuste y se deberían tomar medidas para reducirlo, como ajustar el modelo, recopilar más datos o utilizar técnicas de regularización.

Para este modelo en particular, la diferencia del 3% entre el *F1-score* de entrenamiento y prueba sugiere que no hay sobreajuste significativo.

Optimización de parámetros

Para optimizar los parámetros del modelo, se utilizó *RandomizedSearchCV*, resultando en los siguientes valores:

- *max_depth* = 10, equilibrando la complejidad del modelo y su capacidad de generalización.
- *max_features = None*, permite al modelo utilizar todas las características disponibles, lo que puede mejorar la precisión si las características son relevantes.
- *min_samples_leaf* = 2, enfoque hacia una estructura más flexible del modelo, permitiendo que algunas hojas de los árboles de decisión tengan pocas observaciones.
- *min_samples_split* = 2, indica un enfoque más propenso a sobreajuste, ya que requiere menos observaciones para dividir un nodo en el árbol de decisión.
- *n_estimators:* 10, se utilizaron 10 árboles en el modelo, valor moderado.

En la sección del Apéndice 9.5.3 se explora el *feature importance* del modelo.

Análisis de la matriz de confusión

La matriz de confusión (Figura 4.8) revela un desempeño excelente en la clase "0-10" (99% de precisión) debido al desbalance de clases. A pesar de representar solo el 2% del total de observaciones, el modelo logra un 27% de precisión en la clase "≥ 51". Sin embargo, presenta un bajo rendimiento en las clases intermedias "11-20" y "21-50" (0% de precisión en ambas).

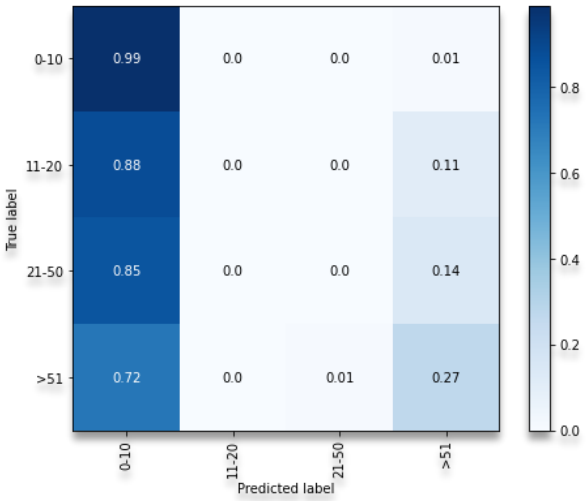


Figura 4.8 Matriz de confusión del modelo Random Forest Classifier con 4 clases

Conclusiones

El *Random Forest Classifier* con cuatro clases demostró ser efectivo para evaluar la regla de negocio actual, mostrando un alto rendimiento general. No obstante, se identifican áreas de mejora en las clases intermedias. Por lo tanto, a continuación, se explora *XGBoost* como modelo alternativo, con el objetivo de superar las limitaciones encontradas y mejorar la precisión, generalización e interpretabilidad en estas clases.

4.2.2 *Modelo 4: XGBoost Classifier con Cuatro Clases*

Rendimiento del modelo

El modelo *XGBoost Classifier* con cuatro clases demuestra un rendimiento superior en comparación al modelo *Random Forest Classifier*. El modelo alcanzó un F1-score de 0.952 en el conjunto de entrenamiento y 0.914 en el conjunto de prueba, representando un incremento del 1.48% y 0.55% respectivamente respecto al modelo *Random Forest Classifier.*

Aunque la diferencia entre el *F1-score* de entrenamiento y prueba aumentó ligeramente de 0.029 a 0.038, no supera el umbral del 5% establecido para considerar un sobreajuste significativo.

Optimización de parámetros

Durante la optimización de hiperparámetros determinaron los siguientes valores:

- *max_depth* = 7, indica una profundidad moderada de los árboles de decisión, permitiendo que el modelo aprenda patrones complejos sin llegar a memorizar datos específicos.
- *n_estimators* = 10, se utilizaron 10 árboles en el modelo, un valor moderado entre las opciones.
- *learning_rate* = 1, indica un aprendizaje agresivo del modelo, lo que puede contribuir al sobreajuste. Sin embargo, el bajo *gap* entre el *F1-score* de entrenamiento y prueba sugiere que el aprendizaje agresivo no ha generado un sobreajuste significativo.

Análisis de la matriz de confusión

El análisis de la matriz de confusión (Figura 4.9) muestra un desempeño superior del *XGBoost Classifier* en comparación con el *Random Forest Classifier.* especialmente en las clases intermedias "11-20" y "21-50". La precisión en estas clases aumentó notablemente, pasando del 0% a 4% y 5% respectivamente. En la clase menos frecuente "≥ 51", la precisión también mejoró significativamente, aumentando del 27% al 36%. La clase mayoritaria "0-10" mantuvo una alta precisión del 99%, lo cual es consistente con el desbalance de clases en el conjunto de datos.

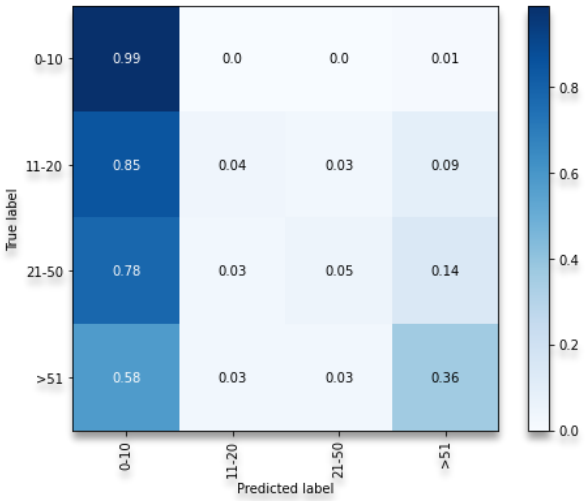


Figura 4.9 Matriz de confusión del modelo XGBoost Classifier con 4 clases

En la sección del Apéndice 9.5.4 se explora el *feature importance* del modelo.

Conclusiones

En resumen, el modelo *XGBoost Classifier* ha demostrado una mejora sustancial sobre el *Random Forest Classifier*, especialmente en clases intermedias y menos frecuentes. Esta mejora se atribuye a la capacidad de *XGBoost* para manejar patrones complejos sin sobreajustarse a los datos.

A pesar de estos avances, se explora una estrategia alternativa de clasificación con menos clases para evaluar si puede mejorar aún más el rendimiento del modelo. Esta decisión busca encontrar un equilibrio entre la precisión y la granularidad de las predicciones, considerando las particularidades de cada clase en el problema de estimación de demanda de productos sin historial de ventas.

4.2.3 *Modelo 5: Random Forest Classifier con Tres Clases*

El modelo 5 adopta una nueva aproximación al clasificar los SKUs en tres clases en lugar de cuatro: "0-10", "11-20" y "≥ 21". Estas clases representan el 93.9%, 2.1% y 3.9% del total de observaciones respectivamente, como se muestra en la Figura 4.10.

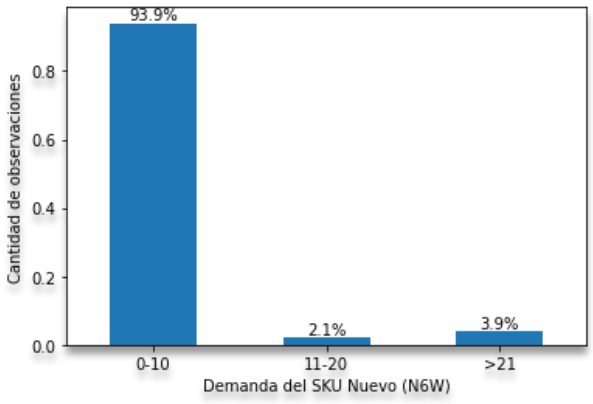


Figura 4.10 Distribución de variable target con tres clases

La simplificación del modelo persigue objetivos que benefician tanto el análisis como el rendimiento del modelo:

1. Simplificación del enfoque: Reducir las clases intermedias disminuye la complejidad del análisis, facilitando la interpretación de resultados y la toma de decisiones estratégicas. Esto permite concentrarse en categorías principales y evitar sesgos por excesiva segmentación.
2. Mejora del rendimiento del modelo: Al agrupar clases, se reduce el ruido en los datos, mejorando la precisión y estabilidad de las predicciones. Menos categorías permiten al modelo enfocarse en patrones más relevantes, optimizando la identificación precisa de SKUs.
3. Reducción de la complejidad del modelo: Menos clases implican un modelo menos complejo, lo que resulta en entrenamientos más rápidos y mejor generalización a nuevos datos. Esto optimiza el proceso analítico y decisional, ahorrando recursos computacionales.
4. Identificación de SKUs clave: La segmentación busca identificar con precisión los SKUs en la clase "≥ 21", que representan productos con alto volumen de ventas. Esto permite flexibilizar las restricciones de inventario, optimizando niveles de *stock*, reduciendo costos y evitando pérdidas por falta de productos demandados.

Rendimiento del modelo

El modelo ha demostrado un sólido rendimiento, logrando un *F1-score* de 0.961 en el conjunto de entrenamiento y 0.919 en el conjunto de prueba. Estos resultados superan a los obtenidos por los modelos de clasificación previos, evidenciando la efectividad de la estrategia de agrupación de clases. El *gap* entre *F1-score* de entrenamiento y prueba (4.6%) no indica sobreajuste significativo, ya que está por debajo del umbral del 5% establecido para este análisis.

Optimización de parámetros

La optimización de los parámetros se realizó utilizando *RandomizedSearchCV*, resultando en:

- *max_depth* = 20.
- *max_features* = *'sqrt'*.
- *min_samples_leaf* = 4.
- *min_samples_split* = 2.
- *n_estimators* = 25.

Esta configuración equilibra la precisión del modelo con la eficiencia computacional, permitiendo obtener resultados confiables con un tiempo de entrenamiento razonable.

Análisis de la matriz de confusión

A pesar de su simplicidad en la segmentación de clases, el modelo no logra mejorar la estimación de la clase intermedia "11-20" y enfrenta desafíos en la estimación de la clase ">21". La matriz de confusión (Figura 4.11) revela un 0% de precisión en la clase "11-20" y un 25% en la clase ">21". Estos resultados indican que, si bien el modelo presenta un buen rendimiento general, se requieren estrategias adicionales para mejorar la precisión en estas dos clases específicas.

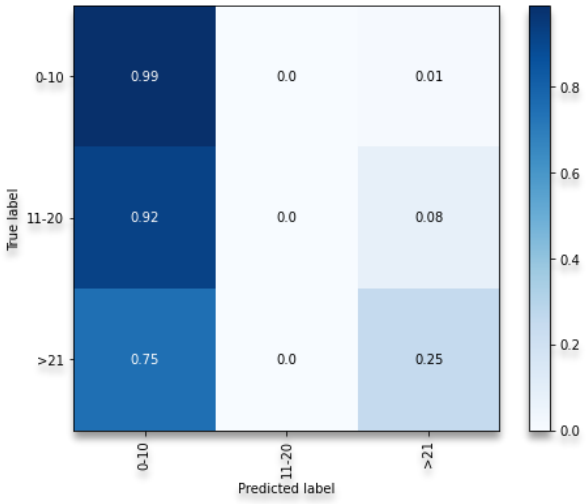


Figura 4.11 Matriz de confusión del modelo random forest classifier con 3 clases

En la sección del Apéndice 9.5.5 se explora el *feature importance* de cada clase.

Conclusiones

Dadas las limitaciones observadas en la estimación de las clases "11-20" y ">21" se ha decidido entrenar un modelo adicional utilizando *XGBoost* con tres clases para abordar estos desafíos. Se espera

que el modelo *XGBoost*, con su capacidad mejorada para aprender patrones complejos, pueda superar estas deficiencias y proporcionar estimaciones más precisas para todas las clases del problema de estimación de demanda de productos sin historial de ventas.

4.2.4 *Modelo 6: XGBoost Classifier con Tres Clases*

El modelo 6 utiliza *XGBoost* para clasificar SKUs en tres clases "0-10", "11-20" y "≥ 21", y busca mejorar el rendimiento del modelo anterior.

Rendimiento del modelo

El modelo *XGBoost* con tres clases alcanza un *F1-score* de 0.998 en el conjunto de entrenamiento, superando al *Random Forest* en este aspecto. Sin embargo, ambos modelos obtienen un *F1-score* de 0.919 en el conjunto de prueba.

La diferencia del 8.6% entre los resultados de entrenamiento y prueba sugiere un sobreajuste moderado en el modelo *XGBoost*, que supera el umbral del 5% pero se mantiene por debajo del 10%, lo que sugiere un caso de sobreajuste menos severo. Esto indica que el modelo ha aprendido patrones específicos de los datos de entrenamiento, limitando su capacidad de generalización a nuevos datos.

Optimización de parámetros

La optimización de hiperparámetros con '*RandomizedSearchCV*' reveló la siguiente configuración óptima para el modelo:

- *max_depth* = 20, el máximo entre las opciones disponibles.
- *n_estimators* = 25.
- *learning_rate* = 0.5, lo que indica un ritmo de aprendizaje moderado.

Si bien la configuración óptima logra un buen rendimiento, la profundidad máxima elevada representa un riesgo potencial de sobreajuste.

Análisis de la matriz de confusión

La matriz de confusión del modelo *XGBoost* (Figura 4.12) muestra mejor desempeño que el *Random Forest* en las tres clases. Destaca con un 1% de precisión en la clase "11-20", un 99% en la clase "0-10" y un 30% en la clase "≥ 21", lo cual es relevante para la toma de decisiones en esta investigación.

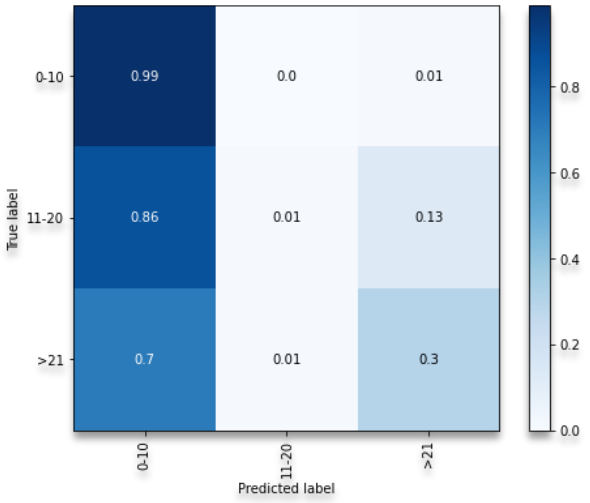


Figura 4.12 Matriz de confusión del modelo XGBoost Classifier con 3 clases

En la sección del Apéndice 9.5.6 se explora el *feature importance* del modelo.

Conclusiones

Basado en el rendimiento prometedor observado, se decidió entrenar un nuevo modelo *XGBoost* con tres clases para ajustar los hiperparámetros y mitigar el sobreajuste detectado. Este enfoque busca mejorar la consistencia entre los resultados del entrenamiento y la prueba, optimizando la capacidad predictiva del modelo para las clases específicas definidas.

4.2.5 *Modelo 7: XGBoost Classifier con Tres Clases – Reducción de Overfitting*

El Modelo 7 es una versión ajustada del clasificador *XGBoost* con tres clases, diseñada para mitigar el sobreajuste y mejorar la generalización del modelo.

Optimización de parámetros

Se utilizó *RandomizedSearchCV* para optimizar los parámetros del modelo.

- *max_depth:* Se probó con valores [7, 9, 11, 13], y se seleccionó *max_depth* = 9.
- *n_estimators:* Se probó con valores [5, 7, 10], y se seleccionó *n_estimators* = 10.
- *learning_rate:* el rango de valores se redujo a [0.4, 0.45, 0.5], y el valor optimizado fue 0.5.

Rendimiento del modelo

El ajuste de hiperparámetros dio como resultado un *F1-score* de 0.921 en datos de prueba, representando mejoras respecto al modelo 6. Sin embargo, el *F1-score* en datos de entrenamiento

experimentó una leve disminución a 0.957. Sin embargo, la brecha entre entrenamiento y prueba se redujo considerablemente a 4%, indicando una mejor capacidad de generalización del modelo.

Análisis de la matriz de confusión

La matriz de confusión (Figura 4.13) revela un rendimiento sobresaliente en la clase "0-10", con casi todas las observaciones clasificadas correctamente. En la clase "≥ 21", el modelo alcanza un 29% de precisión, mostrando un buen desempeño en esta categoría menos frecuente. No obstante, la clase intermedia "11-20" presenta un desempeño bajo con un 0% de precisión, señalando un área de mejora para futuras iteraciones del modelo.

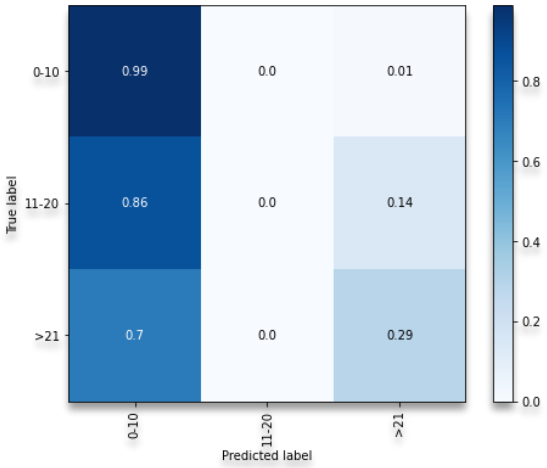


Figura 4.13 Matriz de confusión del modelo XGBoost Classifier con 3 clases

Feature importance

La Figura 4.14 muestra el *feature importance* para la clase "0-10". La variable más relevante es la reputación del vendedor *'green platinum'*, con un peso superior a 0.35. Esto indica que esta *feature* tiene un impacto significativo en la predicción de la clase "0-10". Las demás variables tienen una importancia relativamente menor.

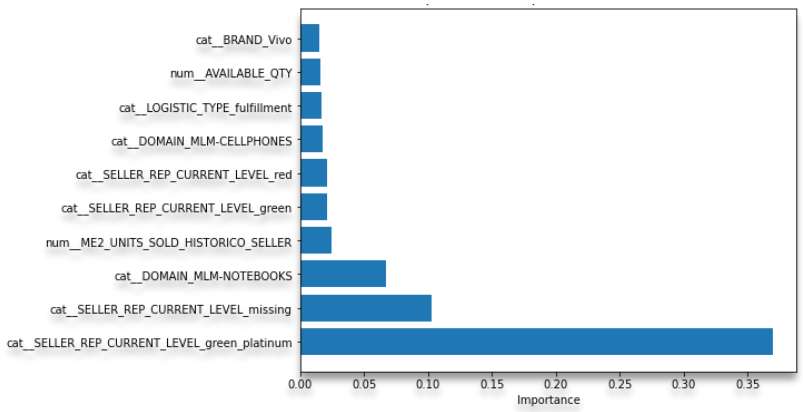


Figura 4.14 Top diez features más relevantes del modelo entrenado para la clase "0-10"

En la Figura 4.15 se presenta el *feature importance* para la clase "11-20". A diferencia de la clase "0-10", las variables en esta clase tienen una importancia individual más baja y y distribuida de manera más equilibrada. Las tres características más importantes son el histórico de ventas del vendedor y su nivel de reputación, especialmente en las categorías *'red'* y *'green'*.

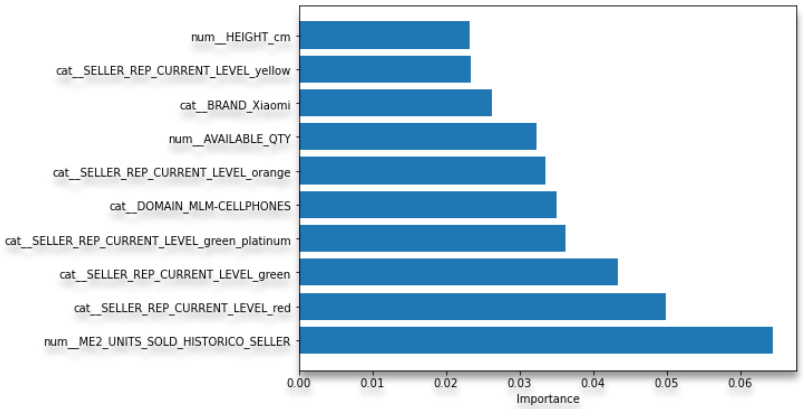


Figura 4.15 Top diez features más relevantes del modelo entrenado para la clase "11-20"

La Figura 4.16 muestra el *feature importance* para la clase "≥ 21. Se observa una marcada diferencia en la importancia de las variables en esta clase. La reputación del vendedor, particularmente en el nivel '*green platinum*', es la variable más relevante, con una importancia relativa superior a 0.40. Le sigue, con un peso significativamente menor, el dominio de *notebooks* en México.

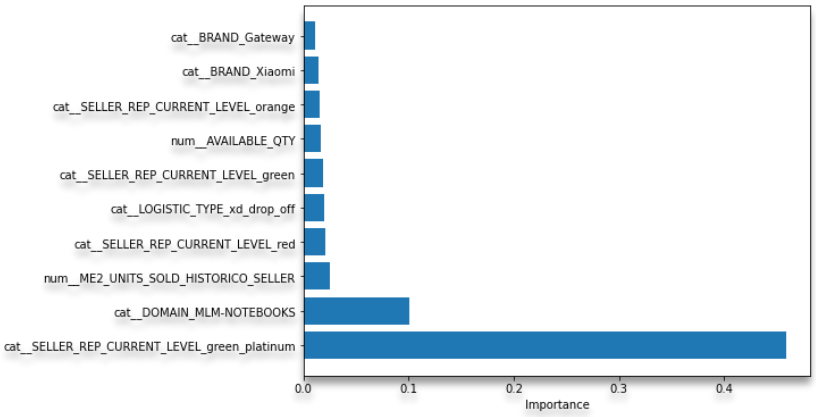


Figura 4.16 Top diez features más relevantes del modelo entrenado para la clase “>21”

Los resultados del análisis del *feature importance* subrayan la importancia crítica de la reputación del vendedor en la determinación de la demanda de un SKU nuevo en las primeras seis semanas de ventas. Este hallazgo respalda la importancia de considerar la reputación del vendedor al desarrollar estrategias de políticas de ingreso de inventario a los centros de *fulfillment* del *marketplace*.

Conclusiones

El presente estudio ha desarrollado un modelo de aprendizaje automático para optimizar la gestión de inventario en un marketplace de comercio electrónico. El modelo final se destaca por su capacidad para reducir el *overfitting* y mejorar los resultados en pruebas, identificando con precisión las clases extremas de demanda: "0-10" y "≥21". Esta precisión permite optimizar la gestión del inventario, ajustando los umbrales de unidades permitidas para cada clase y maximizando el uso del espacio de almacenamiento.

Si bien el modelo no predice con precisión la clase intermedia "11-20", esta limitación se mitiga parcialmente por el hecho de que la mayoría de las observaciones de esta clase se clasifican erróneamente como "0-10" (86%), mientras que una porción mucho menor se clasifica como "≥21" (14%). Esto reduce el riesgo inherente del error de predicción en términos de impacto en costos y espacio de almacenamiento para los centros de *fulfillment* del *marketplace*. Para la clase "11-20", que hoy tiene una política que permite ingresar hasta 20 unidades por SKU, los resultados del modelo podrían reemplazarse por una nueva política que limite, en la mayoría de los casos, el ingreso a un máximo de 10 unidades, minimizando el riesgo de sobreabastecimiento.

Aunque una solución simplista para mejorar la precisión general del modelo sería agrupar las clases "0-10" y "11-20" en una sola categoría (“0-20”), esta opción no es viable debido a las desventajas significativas que conlleva. La pérdida de granularidad y la limitación en la toma de decisiones impedirían

la implementación de políticas diferenciadas de ingreso de inventario, afectando negativamente la optimización del inventario en el segmento "0-10".

Para complementar la evaluación del modelo y abordar la limitación en la predicción de la clase "11-20", se ha realizado un análisis complementario con matrices de probabilidad para cada clase que permite evaluar la calibración del modelo y tomar decisiones basadas en el nivel de confianza deseado.

En definitiva, el modelo desarrollado en este estudio presenta un gran potencial para optimizar las políticas de inventario del *marketplace*, a pesar de la limitación en la predicción de la clase "11-20". Al enfocarse en las fortalezas del modelo y aprovechar el análisis de probabilidades, el *marketplace* puede tomar decisiones más informadas y estratégicas, mejorando la eficiencia, la rentabilidad y la satisfacción de vendedores y compradores. La implementación exitosa de este modelo puede generar un impacto positivo significativo en el negocio, consolidando la posición del *marketplace* en el mercado y contribuyendo a su crecimiento sostenido.

4.3 Selección del Modelo

Durante el desarrollo de esta tesis, se implementaron y evaluaron exhaustivamente distintos modelos de regresión y clasificación para estimar la demanda en las primeras seis semanas de ventas de un producto nuevo sin historial de ventas, basándose en el rendimiento de productos similares en el mercado. El propósito del desarrollo del modelo es enriquecer la toma de decisiones del *marketplace* en sus centros de distribución, optimizando así el uso de los recursos disponibles.

El modelo seleccionado fue un clasificador de tres categorías que emplea el algoritmo *XGBoost*, debido a su desempeño superior. Los resultados del modelo demostraron su capacidad para identificar correctamente la clase correspondiente para observaciones fuera del conjunto de entrenamiento, gracias a un meticuloso proceso de entrenamiento y ajuste de hiperparámetros.

El modelo alcanzó un *F1-score* de 0.921, superando a los modelos de regresión y clasificación. Este éxito se debe a su capacidad para capturar relaciones no lineales y entender interacciones complejas entre las variables predictoras.

El *output* del modelo se presenta como probabilidades de pertenecer a cada categoría de demanda en las primeras seis semanas de ventas. Estas probabilidades ofrecen una perspectiva detallada sobre la confianza del modelo en sus predicciones, lo cual resulta muy útil no solo para tomar decisiones

informadas en la gestión de la demanda, sino también para optimizar la asignación de recursos de los centros de distribución del *marketplace* y planificar estrategias basadas en la probabilidad de que se concreten ciertos niveles de demanda.

Es importante destacar que, si bien el modelo *XGBoost* demostró excelentes resultados en entrenamiento y prueba, no puede considerarse como una predicción infalible. La precisión de los resultados depende de la calidad y cantidad de datos utilizados en el entrenamiento, así como de la selección apropiada de variables significativas. El mantenimiento constante y la reevaluación periódica del modelo son esenciales para asegurar su desempeño a lo largo del tiempo, ya que las condiciones y los factores que afectan la demanda pueden variar con el tiempo. Esta cuestión es especialmente relevante en el contexto de productos tecnológicos, donde la demanda tiende a fluctuar rápidamente a medida que son reemplazados por productos más modernos.

En la siguiente sección, se presentan las matrices de probabilidad para cada clase del modelo, complementando y ampliando las métricas previamente evaluadas, como el *F1-score* y las matrices de confusión.

4.3.1 *Feature Importance de la Clase “>21” para el Modelo Seleccionado*

A continuación, se analiza la clase ">21" y su relación con las variables más relevantes del modelo para determinar la correlación entre la demanda adicional y estas variables.

Nivel de reputación del vendedor

El *boxplot* de la Figura 4.17 compara la demanda de SKUs nuevos en la clase ">21" en función del nivel de reputación del vendedor y la Figura 4.18 detalla las principales medidas estadísticas.

Se observa que la categoría más relevante para identificar la clase ">21", vendedores con reputación *'green platinum'*, presenta un percentil 75 de 86 unidades demandadas en las primeras seis semanas. Esto confirma que, a mejor reputación del vendedor, mayor probabilidad de que sus productos nuevos tengan alta demanda, incluso sin historial de ventas.

La cuarta variable más relevante para la clase “≥ 21”, vendedores con reputación *'red'*, muestra una mediana de 46 unidades demandadas en las primeras seis semanas. Un dato interesante es que, a pesar de ser el peor nivel de reputación, esta categoría no coincide con la mediana más baja, que corresponde a vendedores con reputación *'light green'*, con una mediana de 38 unidades demandadas.

Las categorías *'green'* y *'orange'*, ubicadas en el sexto y octavo lugar en términos de relevancia para identificar la clase "≥ 21", respectivamente, también presentan diferencias. La categoría *'green'* tiene un percentil 75 de 75 unidades demandadas (la segunda categoría con percentil 75 más alto) y *'orange'* un percentil 75 de 63.

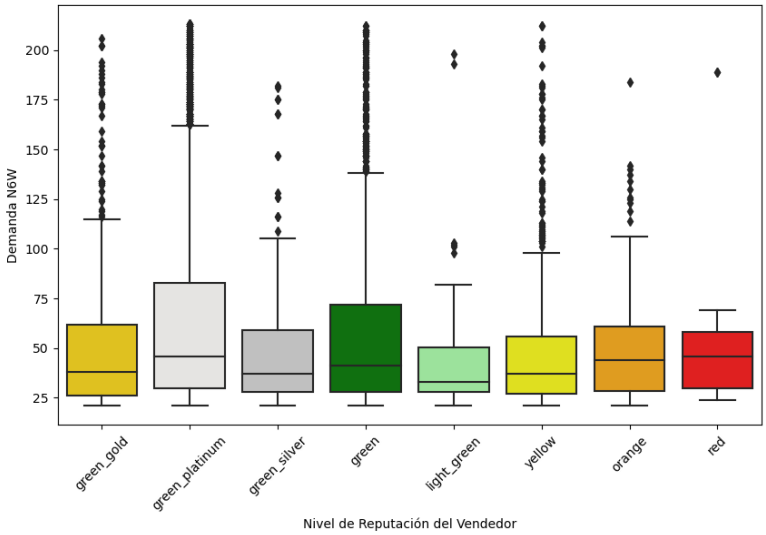


Figura 4.17 Boxplot de la demanda de SKUs nuevos según reputación del vendedor para la clase '>21' (incluye hasta Q 99.5 del dataset)

	Reputación del vendedor							
	Green gold	*Green platinum*	*Green silver*	*Green*	*Light green*	*Yellow*	*Orange*	*Red*
Q25	28	32	30	30	28	29	30	30
Q50	41	49	39	44	38	39	46	46
Q75	65	86	59	75	53	57	63	58

Figura 4.18 Medidas estadísticas de la demanda por reputación del vendedor para la clase '>21'

Dominio

El dominio de la publicación es la segunda variable más relevante para identificar SKUs con alta demanda (clase ">21"), en particular el dominio "MLM – *Notebooks*".

El *boxplot* de la Figura 4.19 y medidas estadísticas de la Figura 4.20 muestra que la demanda varía ligeramente entre dominios, pero no hay diferencias estadísticamente significativas a nivel país (Brasil/México) ni producto (*notebooks/cellphones*).

A nivel país, la mediana de la demanda varía entre 41 y 49 unidades, sin patrones claros que indiquen una mayor o menor demanda en uno u otro país. A nivel producto, la mediana de la demanda para *notebooks* es levemente mayor que la de *cellphones* (en promedio 48 unidades vs. 43), pero la diferencia no es estadísticamente significativa.

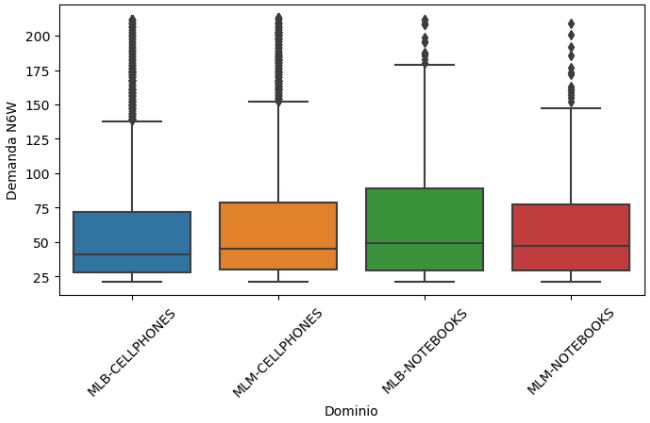


Figura 4.19 Boxplot de la demanda de SKUs nuevos según dominio para la clase ">21"

	Dominio			
	MLB – Cellphones	*MLM – Cellphones*	*MLB - Notebooks*	*MLM – Notebooks*
Q25	28	30	29	29
Q50	41	45	49	47
Q75	72	79	89	77

Figura 4.20 Medidas estadísticas de la demanda por dominio para la clase '>21'

Histórico de Ventas del Vendedor

El histórico de ventas del vendedor es la tercera variable más relevante para identificar la clase "≥ 21". Sin embargo, el diagrama de dispersión (Figura 4.21) no muestra una relación clara entre las ventas anteriores y la demanda actual, indicando que no hay un patrón definido.

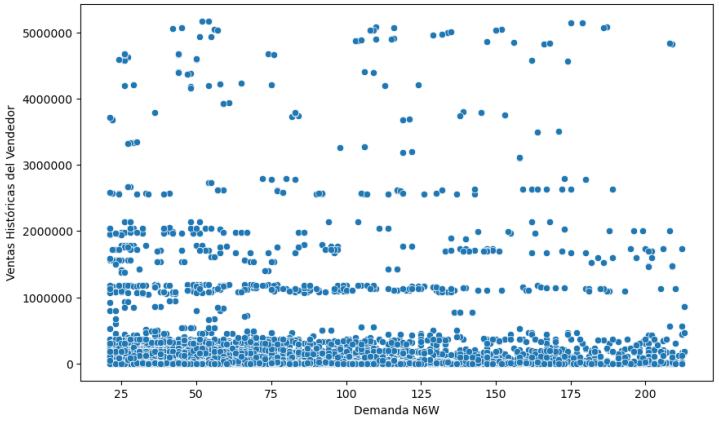


Figura 4.21 Demanda de SKUs nuevos según histórico de ventas del vendedor para la clase ">21"

Logistic Type

El tipo de logística de la publicación es el quinto factor más relevante para identificar SKUs con alta demanda (clase "≥ 21"), en particular, la logística *'XD drop off'*. A continuación, se analiza la relación entre ambas variables, utilizando *boxplot* (Figura 4.22) y medidas estadísticas (Figura 4.23).

Si bien las diferencias observadas no son estadísticamente significativas, *'fulfillment'* presenta consistentemente mejores resultados en términos de volumen de demanda para SKUs nuevos sin historial. Por un lado, presenta la mediana más alta con 54 unidades, mientras que las restantes logísticas oscila entre 40 y 46. También destaca en el percentil 75, con 107 unidades, mientras que las demás oscilan entre 68 y 79 unidades.

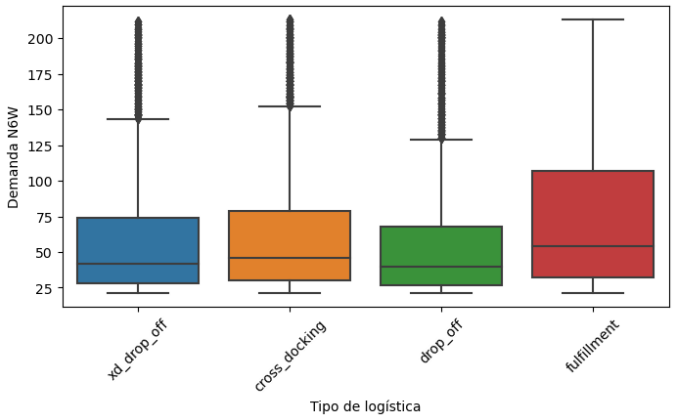


Figura 4.22 Boxplot de la demanda de SKUs nuevos según tipo de logística para la clase '>21'

	Tipo de logística			
	XD Drop off	*Cross Docking*	*Drop-Off*	*Fulfillment*
Q25	28	30	27	32
Q50	42	46	40	54
Q75	74	79	68	107

Figura 4.23 Medidas estadísticas de la demanda por tipo de logística para la clase '>21'

Unidades Disponibles a la Venta

Las unidades disponibles a la venta al momento de la creación de la publicación es la séptima más relevante para identificar SKUs con alta demanda (clase ">21"). El diagrama de dispersión (Figura 4.24) muestra que no hay una relación clara entre las unidades disponibles y la demanda del SKU.

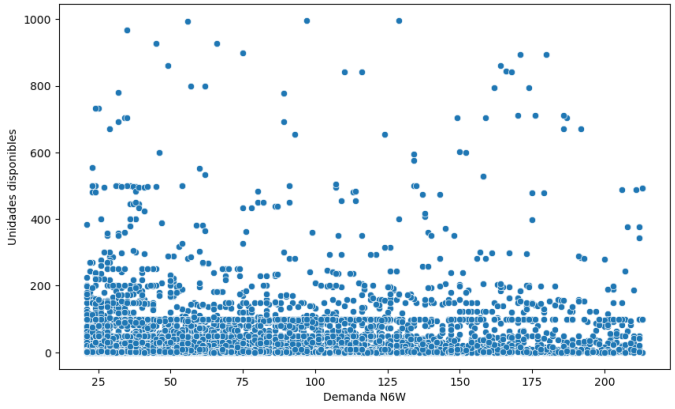


Figura 4.24 Demanda de SKUs nuevos según unidades disponibles a la venta para la clase ">21"

Conclusiones

El análisis de la clase "≥21" revela tendencias asociadas al nivel de reputación del vendedor, dominio del producto y tipo de logística. Sin embargo, estos patrones no son concluyentes para diseñar políticas de ingreso diferenciado dentro de la clase "≥21". Por ello, se analizaron las matrices de probabilidad del modelo para determinar la adopción de diferentes políticas de ingreso basadas en distintos niveles de confianza. Estos hallazgos son fundamentales para desarrollar estrategias comerciales precisas y efectivas, optimizando la gestión de inventario en *fulfillment* y maximizando las oportunidades de venta en el *marketplace*.

4.3.2 *Matriz de Probabilidad del Modelo Seleccionado*

Las matrices de probabilidades proporcionan una visión detallada no solo de la precisión general del modelo, sino también de la confianza que este tiene en sus predicciones para diferentes clases reales. A diferencia de métricas como la precisión o el *F1-score*, que solo consideran clasificaciones correctas o incorrectas, las matrices de probabilidad brindan información detallada sobre la confianza que el modelo tiene en cada predicción.

En el contexto de este estudio, donde se busca optimizar la gestión de SKUs nuevos, las matrices de probabilidad son herramientas invaluables para tomar decisiones informadas basadas en el nivel de tolerancia al riesgo del *marketplace* en cuestión.

Confección de Matrices de Probabilidad

La matriz de probabilidades final refleja la probabilidad de que una observación pertenezca a una clase real específica, dado un rango de probabilidad predicho por el modelo. Los pasos son:

1. Identificación de clases reales: Las clases consideradas son "0-10", "11-20" y ">21" unidades demandadas por SKU en las primeras seis semanas.
2. Construcción de tablas de contingencia: Se crea una tabla para cada clase real que muestra la frecuencia de observaciones predichas por el modelo dentro de diferentes rangos de probabilidad. Las columnas que representan los rangos de probabilidad del modelo (por ejemplo, "0-20%", "20-40%", etc.) y filas que representan las clases predichas.
3. Cálculo de frecuencias: Las frecuencias se calculan comparando las predicciones del modelo con las clases reales conocidas.

4. Normalización de tablas: Para obtener probabilidades en lugar de frecuencias absolutas, se normalizan las tablas dividiendo cada valor por el total de su columna correspondiente.
5. Interpretación: El análisis de estas matrices revela los rangos de probabilidad donde el modelo muestra mayor confianza en sus clasificaciones, facilitando decisiones más seguras. Además, permite comparar la distribución de predicciones dentro de cada rango con la distribución general de los datos para cada clase.

Matriz de Probabilidad: Clase "0-10"

En la Figura 4.26 se observa la matriz de probabilidades para la clase "0-10". Se destaca un alto grado de confianza en los rangos superiores ("40-60", "60-80" y "80-100%"), donde el modelo predice correctamente el 100% de las observaciones como pertenecientes a la clase "0-10". Esto indica una certeza significativa de que la demanda real de estos SKUs será inferior a 10 unidades.

Incluso en el rango "20-40%", el modelo muestra un desempeño sólido, clasificando correctamente el 96% de las observaciones en la clase "0-10", superando la proporción esperada de observaciones para esta clase en la distribución general de datos (93%). Estos resultados subrayan la capacidad del modelo para ofrecer predicciones precisas y útiles, superando una simple estimación basada en la distribución de datos.

	Matriz de probabilidades - clase "0-10"					
	Prob 0-20	Prob 20-40	Prob 40-60	Prob 60-80	Prob 80-100	All
0-10	9.734	12.932	13.463	13.445	13.484	**63.058**
11-20	1.288	296	65	4	2	**1.655**
>21	2.487	280	53	8	0	**2.828**
All	**13.509**	**13.508**	**13.581**	**13.457**	**13.486**	**67.541**

Figura 4.25 matriz de probabilidades para la clase "0-10" en cantidad de observaciones

	Matriz de probabilidades - clase "0-10"					
	Prob 0-20	Prob 20-40	Prob 40-60	Prob 60-80	Prob 80-100	All
0-10	0.72	0.96	1.00	1.00	1.00	**0.93**
11-20	0.10	0.02	0	0	0	**0.02**
>21	0.18	0.02	0	0	0	**0.04**
All	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**

Figura 4.26 matriz de probabilidades para la clase "0-10" en porcentaje

Matriz de Probabilidad: Clase "11-20"

La Figura 4.28 muestra la matriz de probabilidades para la clase "11-20". Aunque el modelo supera la distribución normal de las observaciones en los últimos dos quintiles ("60-80" y "80-100%") con un 3% y 9% de predicciones en esta clase, respectivamente, comparado con el 2% de la distribución original, su desempeño global en esta categoría es variable.

	Matriz de probabilidades - clase "11-20"					
	Prob 0-20	Prob 20-40	Prob 40-60	Prob 60-80	Prob 80-100	All
0-10	13.507	13.493	13.376	12.682	10.000	**63.058**
11-20	2	5	59	339	1.250	**1.655**
>21	0	10	73	490	2.255	**2.828**
All	**13.509**	**13.508**	**13.508**	**13.511**	**13.505**	**67.541**

Figura 4.27 matriz de probabilidades para la clase "11-20" en cantidad de observaciones

	Matriz de probabilidades - clase "11-20"					
	Prob 0-20	Prob 20-40	Prob 40-60	Prob 60-80	Prob 80-100	All
0-10	1.00	1.00	0.99	0.94	0.74	**0.93**
11-20	0	0	0	0.03	0.09	**0.02**
>21	0	0	0.01	0.04	0.17	**0.04**
All	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**

Figura 4.28 matriz de probabilidades para la clase "11-20" en porcentaje

Es importante tener en cuenta que la clase "11-20" representa un porcentaje bajo de las observaciones totales. El verdadero valor del modelo radica en su capacidad para identificar con precisión las categorías de demanda alta (">21") y baja ("0-10") para permitir al *marketplace* ajustar las políticas de inventario vigentes que limitan el ingreso de inventario a 20 unidades por SKU.

Matriz de Probabilidad: Clase "≥ 21"

La matriz de probabilidades para la clase "≥ 21" (Figura 4.30) revela resultados alentadores, especialmente en el quintil "80-100". En este rango, el modelo identifica correctamente el 18% de los SKUs como pertenecientes a la categoría de demanda alta ("≥ 21"). Esto representa un aumento significativo en comparación con la distribución normal de los datos, donde solo el 4% de las observaciones se clasifican en esta categoría, un aumento de 3.5X.

	Matriz de probabilidades - clase ">21"					
	Prob 0-20	Prob 20-40	Prob 40-60	Prob 60-80	Prob 80-100	All
0-10	16.089	10.919	13.373	12.858	9.819	**63.058**
11-20	5	5	65	351	1.229	**1.655**
>21	4	4	61	299	2.460	**2.828**
All	**16.098**	**10.928**	**13.499**	**13.508**	**13.508**	**67.541**

Figura 4.29 matriz de probabilidades para la clase ">21" en cantidad de observaciones

	Matriz de probabilidades - clase ">21"					
	Prob 0-20	Prob 20-40	Prob 40-60	Prob 60-80	Prob 80-100	All
0-10	1.00	1.00	0.99	0.95	0.73	**0.93**
11-20	0	0	0	0.03	0.09	**0.02**
>21	0	0	0	0.02	0.18	**0.04**
All	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**

Figura 4.30 matriz de probabilidades para la clase ">21" en porcentaje

Conclusiones

Este análisis detallado de las predicciones del modelo, segmentadas por la probabilidad de ocurrencia, proporciona una herramienta valiosa para tomar decisiones más precisas y fundamentadas. En lugar de depender de reglas genéricas que se aplican uniformemente a todas las situaciones, este enfoque permite evaluar cada predicción en función de su nivel específico de confianza. Adaptando las decisiones a la fiabilidad intrínseca de cada predicción, es posible asignar recursos, implementar estrategias y tomar medidas basadas en la certeza que el modelo muestra en cada caso particular.

5 Limitaciones del Modelo

En esta sección, se examinan las limitaciones de los modelos entrenados para predecir la demanda de productos nuevos sin historial de ventas. Se resalta la dificultad de aplicar estos modelos en productos poco estandarizados, las complejidades para anticipar eventos disruptivos con características diferentes a los datos de entrenamiento, y se subraya la necesidad de una aproximación dinámica para mejorar la precisión de las predicciones.

5.1 Tipología de Productos

Una limitante importante del modelo radica en la tipología de los productos analizados. En esta tesis, se ha enfocado en productos altamente estandarizados, específicamente en las categorías de *notebooks* y celulares. Estos productos se caracterizan por su uniformidad en especificaciones técnicas y características medibles, lo que facilita la identificación y análisis de patrones por parte del modelo.

Sin embargo, al intentar aplicar este enfoque a productos menos uniformes, como los de diseño y decoración para el hogar, se enfrentan desafíos significativos. Estos productos presentan una gran diversidad en estilos, materiales, tamaños y colores, lo que dificulta la identificación de patrones coherentes y tendencias claras. La subjetividad y la variabilidad contextual de los atributos de estos productos hacen que la extrapolación de los resultados del modelo sea poco fiable.

Aunque el *marketplace* analizado permite la inclusión de productos de diseño y decoración en sus centros de *fulfillment*, predecir su rendimiento de ventas cuando no tienen historial es especialmente complicado. La falta de datos históricos consistentes y comparables para estos productos poco estandarizados limita la eficacia del modelo.

Para abordar esta limitación, sería necesario explorar enfoques alternativos que puedan manejar la heterogeneidad de los datos, como técnicas de análisis más avanzadas y la integración de fuentes adicionales de información que enriquezcan el modelo y mejoren su capacidad predictiva en el contexto de productos poco estandarizados.

5.2 Anticipación de Eventos Disruptivos

Dentro del espectro de modelos de aprendizaje automático, el *XGBoost* seleccionado como la base para esta investigación se destaca por sus capacidades. Sin embargo, se encuentra enmarcado en un contexto que presenta desafíos notables en relación con la anticipación de cambios drásticos en el comportamiento de los usuarios, tales como los ocasionados por pandemias u otros eventos disruptivos. Los modelos de este tipo, incluido *XGBoost*, fundamentan sus predicciones en el análisis de datos históricos y patrones pasados, lo cual implica que su habilidad para prever eventos excepcionales, que no han sido contemplados en su conjunto de entrenamiento, está intrínsecamente limitada. Este aspecto plantea una preocupación relevante en el contexto del *marketplace*, donde la toma de decisiones precisa y eficiente es esencial.

La limitación mencionada presenta implicaciones directas para el *marketplace*, ya que la falta de consideración de escenarios extremos en el proceso de entrenamiento del modelo puede llevar a decisiones inapropiadas. Por ejemplo, podrían surgir situaciones en las que el modelo recomiende a los vendedores mantener un exceso de inventario de un producto cuyo ciclo de vida es efímero y está sujeto a cambios de tendencia con rapidez. El resultado sería una distribución desbalanceada entre la oferta y la demanda, lo que afectaría negativamente la utilización óptima del espacio de *fulfillment*.

Esta problemática coincide con las observaciones de académicos en el campo. En el artículo '*Overfitting and Undercomputing in Machine Learning*'[58] de Dietterich, se subraya la importancia de la adaptabilidad de los modelos de aprendizaje automático a eventos excepcionales y su susceptibilidad al sobreajuste cuando se enfrentan a situaciones no contempladas en los datos históricos.

A fin de abordar este desafío, es esencial la combinación los modelos predictivos con la intuición y la experiencia humanas. La colaboración entre la capacidad analítica de los modelos y la visión holística de los expertos puede ayudar a contrarrestar las limitaciones inherentes a la incapacidad de anticipar eventos disruptivos.

5.3 Información Estática

La limitación principal del modelo reside en su carácter estático, dado que se entrenó con datos de un momento puntual. Esto implica que las recomendaciones no reflejan las fluctuaciones en las tendencias y patrones de demanda, que son particularmente volátiles en la vertical de electrónica de consumo, donde los productos se actualizan con frecuencia y las líneas antiguas pueden perder rápidamente su demanda. En este entorno, si se introduce un producto nuevo que no comparte

características con productos anteriores en el conjunto de datos, como la marca, el vendedor o la línea, el modelo puede enfrentar dificultades para generar predicciones precisas.

Para superar esta limitación, se propone una aproximación dinámica en la que el modelo pueda reentrenarse periódicamente con información actualizada sobre la demanda de productos nuevos en el *marketplace*. Esto permitiría capturar las tendencias emergentes y los cambios en el universo de productos disponibles o en sus características. Esta adaptabilidad dinámica del modelo facilitaría ajustes continuos en sus predicciones en función de las variaciones del mercado, ofreciendo recomendaciones de inventario más precisas y actualizadas, y contribuyendo así a una gestión de inventario más eficiente y adaptable en el *marketplace*.

6 Recomendaciones de Gestión

En esta sección, se presentan tres recomendaciones basadas en los resultados del modelo para mejorar la gestión de inventario en la plataforma. Estas recomendaciones implican ajustar, mantener o aumentar los límites de inventario para cada segmento de productos nuevos según las predicciones de demanda. Se espera que estas medidas reduzcan los costos de almacenamiento, mejoren la agilidad operativa y aumenten las ventas. Se estima una posible reducción del 10% en el volumen de unidades almacenadas, lo que contribuiría a mejorar la calidad del inventario y la eficiencia operativa de la plataforma.

Con relación a cómo se debe implementar el resultado de este proyecto de investigación por parte del equipo de gestión de inventario del marketplace para mejorar su proceso de toma de decisiones, se sugiere integrar el modelo en la interfaz de gestión de los vendedores, donde estos tienen visibilidad de su inventario en los centros de fulfillment del marketplace. El objetivo es reemplazar el valor fijo actual de 20 unidades permitidas en México y Brasil por las recomendaciones generadas por el modelo *XGBoost* de tres clases optimizado, que devolverá para cada SKU consultado una de las tres clases como sugerencia de ingreso de inventario a fulfillment: “0-10 unidades”, “11-20 unidades” o “21-40 unidades”.

Esto implica que en ciertos productos se sugiere reducir el límite de unidades permitidas por SKU a 10 unidades, en otros se recomienda mantener la regla actual de 20 unidades, mientras que en ciertos casos se propone permitir un mayor ingreso de hasta 40 unidades para maximizar las ventas en la plataforma.

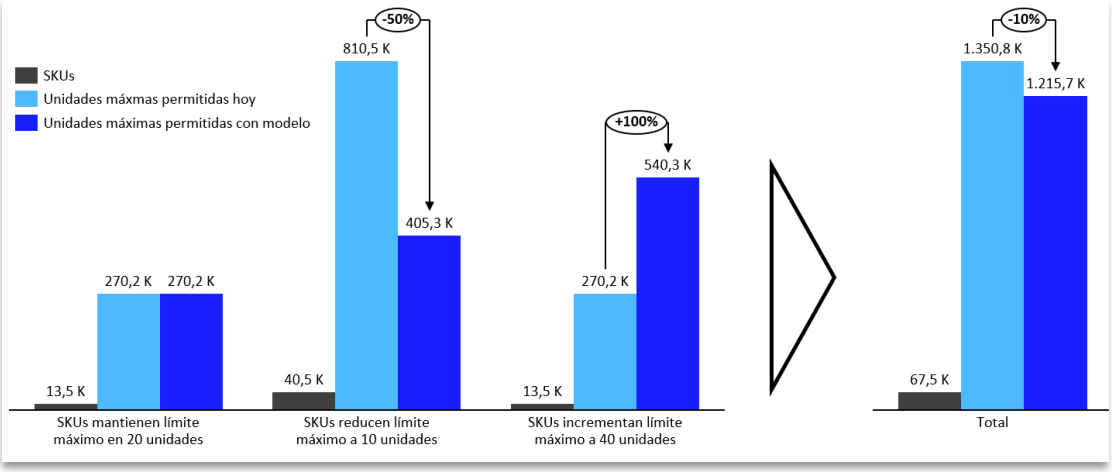


Figura.6.1 Impacto de las recomendaciones en el volumen potencial a almacenar en fulfillment de SKUs nuevos

<u>Recomendación 1:</u> Reducir el límite máximo a 10 unidades para los SKUs nuevos cuya probabilidad predicha de demanda de "0-10 unidades" supere el percentil 40.

Esta medida impactaría al 60% de las observaciones del *dataset* y conllevaría beneficios tanto para el *marketplace* como para los vendedores, gracias a una gestión más eficiente del inventario y a la reducción de los niveles de *stock*.

Desde la perspectiva del *marketplace*, se generan ahorros mediante:

- **Reducción de costos de almacenamiento:** al contar con menos *stock* de productos con baja rotación, se reduce la necesidad de espacio de almacenamiento, lo que resulta en ahorros en alquiler de centros de distribución, costos de mantenimiento y gastos operativos de manejo y almacenamiento.
- **Menor riesgo de daño o pérdida:** al tener niveles de *stock* más bajos en ciertos SKUs, se disminuye el riesgo de daño o pérdida.
- **Mayor agilidad operativa:** la reducción de inventario permite que los centros de distribución operen de manera más ágil, preparando pedidos de forma más rápida y efectiva.

Para los vendedores, se obtienen eficiencias al mantener niveles de *stock* más bajos en los almacenes del *marketplace*:

- **Menores costos de capital inmovilizado:** al reducir el capital invertido en inventario, se libera capital para otras inversiones u operaciones y se reducen posibles costos de financiamiento.
- **Reducción de obsolescencia**: en el caso de productos tecnológicos, como celulares y computadoras, tener niveles más bajos de *stock* cuando se espera baja rotación minimiza el riesgo de obsolescencia.

Como se observa en la Figura 6.1, la implementación de esta recomendación podría resultar en una reducción de hasta el 50% en la cantidad de unidades permitidas para ingresar a *fulfillment* para 40.500 SKUs nuevos de *notebooks* y celulares en México y Brasil. Bajo la política actual del *marketplace*, estos productos podrían ingresar un total de 810.500 unidades en los centros de *fulfillment*. Sin embargo, al aplicar la recomendación, el número total de unidades permitidas se reduciría a 405.300, optimizando así el uso del espacio y los recursos en los centros de distribución.

<u>Recomendación 2:</u> Aumentar el límite máximo a 40 unidades para los SKUs nuevos cuya probabilidad predicha de demanda de ">21 unidades" supere el percentil 80.

Esta medida aplicaría al 20% de las observaciones del *dataset*, lo que brinda ventajas tanto al *marketplace* como a los vendedores.

Desde la perspectiva del *marketplace*, se obtienen ventajas como:

- **Aprovechamiento de la infraestructura logística:** permitir un mayor ingreso de *stock* optimiza el uso de la infraestructura logística del *marketplace*, mejorando la eficiencia operativa.
- **Incremento en comisiones por venta:** al aumentar el inventario de productos con alta rotación, se fomentan más ventas y, por ende, se incrementan las comisiones e ingresos.
- **Ampliación de surtido y variedad:** al cambiar la política de ingreso de *stock* máximo, se incentiva a más vendedores a utilizar los servicios de *fulfillment*, enriqueciendo la oferta de productos y la experiencia del usuario.

Los vendedores también se benefician de incrementar el umbral máximo de *stock* a ingresar:

- **Reducción de costos de envío:** al consolidar los envíos de *stock*, los vendedores obtienen mejores costos de reabastecimiento y pueden ofrecer precios más competitivos.
- **Mayor visibilidad:** los productos almacenados en los centros de *fulfillment* reciben mayor visibilidad, lo que se traduce en mayores ventas y mejores condiciones de entrega para los compradores.

La Figura 6.1 muestra que implementar esta recomendación podría duplicar, o aumentar en un 100%, la cantidad de unidades permitidas para ingresar a *fulfillment* para 13.500 SKUs nuevos de *notebooks* y celulares en México y Brasil. Según la política actual del marketplace, se permitiría el ingreso de hasta 270.200 unidades en total para estos productos en los centros de *fulfillment*. No obstante, con la nueva recomendación, el límite de unidades aumentaría a 540.300. Este ajuste en la política de inventario abriría nuevas oportunidades de venta y ganancia tanto para el *marketplace* como para los vendedores.

<u>Recomendación 3:</u> Mantener el límite máximo de 20 unidades para las restantes casuísticas de SKUs nuevos.

Esta medida impactaría al restante 20% de las observaciones del *dataset*, para los cuales se propone mantener la regla actual del *marketplace*. Esta decisión se basa en la expectativa de que la demanda de estos productos en las primeras seis semanas sea moderada, no justificando ni un aumento ni una disminución en el límite de unidades permitidas.

Los beneficios de mantener el límite actual son:

- **Equilibrio entre oferta y demanda:** al mantener el límite de 20 unidades, se busca un equilibrio entre la oferta disponible y la demanda esperada, evitando tanto el exceso de inventario como la escasez de productos.
- **Optimización del uso del espacio:** al limitar la cantidad de *stock* para estos SKUs, se asegura un uso eficiente del espacio de almacenamiento, permitiendo alojar una mayor variedad de productos en los centros de *fulfillment*.
- **Minimización del riesgo financiero:** al no incrementar el límite máximo de unidades, se reduce el riesgo financiero asociado al capital inmovilizado en inventario, especialmente para aquellos productos con una demanda incierta o fluctuante.

6.1 Impacto Estimado para el Negocio

Al evaluar las tres recomendaciones en conjunto, se estima una reducción potencial de hasta el 10% en el número de unidades que el *marketplace* almacenaría en sus centros de distribución (Figura 6.1). Esto implica pasar de 1.350.800 unidades de SKUs nuevos de *notebooks* y celulares que podrían ingresar bajo la política actual a un total de 1.215.700 unidades.

Sin embargo, más allá de la reducción en el volumen de unidades almacenadas, es esencial considerar la mejora en la calidad de esas unidades en términos de su potencial de rotación. Al alinear el inventario con las necesidades reales de los clientes, se maximiza la probabilidad de que los productos sean vendidos antes de volverse obsoletos, lo que mitigaría el riesgo de pérdidas debido a productos no vendidos y fortalecería la posición competitiva del *marketplace*.

Esta optimización en la gestión del inventario conlleva un uso más eficiente del espacio y los recursos disponibles, lo que se traduce en un aumento de las ventas en la plataforma y una gestión más eficaz del inventario. Además, la mejora en la calidad del inventario puede influir positivamente en la experiencia del cliente, al ofrecer una variedad de productos más relevante y actualizada. Esto puede conducir a una mayor fidelidad del cliente y a un aumento en las recomendaciones positivas, impulsando así el crecimiento y la reputación del *marketplace*.

7 Conclusión

Esta sección destaca el éxito del modelo XGBoost de tres clases en pronosticar la demanda de productos nuevos en el principal marketplace de la región, lo que facilita una gestión de inventario más precisa. Sin embargo, el modelo tiene limitaciones, como su enfoque en productos estandarizados, la dificultad para anticipar eventos disruptivos y su dependencia de datos estáticos. Se sugieren recomendaciones para ajustar el límite de unidades en fulfillment, buscando un equilibrio entre oferta y demanda. Para mantener la eficacia del modelo, es esencial su reevaluación periódica ante cambios en el mercado.

El objetivo principal de esta tesis fue desarrollar un modelo de *machine learning* para pronosticar la demanda de productos nuevos, basándose en el desempeño inicial de productos similares. Este análisis se fundamentó en datos proporcionados por el *marketplace* más importante de América Latina, enfocándose en las ventas históricas en Brasil y México de productos estandarizados, como celulares y *notebooks*.

A través de la exploración, se identificaron patrones en los datos que permitieron comprender qué información debía alimentar al modelo. El modelo construido consideró diferentes tipos de variables, como la marca, el modelo, el precio y las especificaciones técnicas, además del impacto de la reputación del vendedor y su historial de ventas.

Se experimentó con diversos enfoques de modelado, técnicas de *feature engineering*, algoritmos de aprendizaje supervisado y configuración de hiperparámetros. Se aplicó una estrategia de validación cruzada que combinó tanto el *k-fold* como la división de entrenamiento-validación-prueba para evitar el sobreajuste.

Inicialmente, se experimentó con modelos de regresión, los cuales no arrojaron resultados satisfactorios. Posteriormente, se exploraron modelos de clasificación con cuatro y tres clases, obteniendo mejores resultados con este último enfoque. El modelo final, un clasificador *XGBoost* con tres clases, demostró ser capaz de predecir la demanda de productos nuevos en rangos de "0-10 unidades", "11-20 unidades" y ">21 unidades" durante sus primeras seis semanas de ventas en el *marketplace*, con un *F1-score* de 0.92 en datos de prueba.

Para asesorar al *marketplace* sobre la modificación de su política de inventario, se realizaron análisis de las matrices de probabilidad generadas por el modelo seleccionado. Estos análisis permitieron formular recomendaciones que minimizan el riesgo y ofrecen una alternativa a la práctica actual de definir arbitrariamente la cantidad de inventario permitida para productos nuevos en *fulfillment*. Al basar estas decisiones en datos, se puede optimizar la política de inventario de manera más informada y precisa.

En primer lugar, se propone que el *marketplace* reduzca el límite máximo actual de 20 unidades a 10 unidades para los SKUs nuevos cuya probabilidad predicha de demanda de "0-10 unidades" supere el percentil 40. La implementación de esta recomendación podría resultar en una reducción de hasta el 50% en la cantidad de unidades permitidas para ingresar a *fulfillment* para 40,500 SKUs nuevos de *notebooks* y celulares en México y Brasil. Bajo la política actual de la plataforma, estos productos podrían ingresar un total de 810,500 unidades en los centros de *fulfillment*. Sin embargo, al aplicar la recomendación, el número total de unidades permitidas se reduciría a 405,300, optimizando así el uso del espacio y los recursos en los centros de distribución.

Por otro lado, se sugiere aumentar el límite máximo actual de 20 unidades a 40 unidades para los SKUs nuevos cuya probabilidad predicha de demanda de ">21 unidades" supere el percentil 80. La implementación de esta recomendación podría duplicar la cantidad de unidades permitidas para ingresar a *fulfillment* para 13,500 SKUs nuevos de *notebooks* y celulares en México y Brasil. Según la política actual del *marketplace*, se permitiría el ingreso de hasta 270,200 unidades en total para estos productos en los centros de *fulfillment*. No obstante, con la nueva recomendación, el límite de unidades aumentaría a 540,300. Este ajuste en la política de inventario abriría nuevas oportunidades de venta y ganancia tanto para el *marketplace* como para los vendedores.

Por último, se recomienda mantener el límite máximo de 20 unidades para las restantes casuísticas de SKUs nuevos. Esta decisión se basa en la expectativa de que la demanda de estos productos en las primeras seis semanas sea moderada, no justificando ni un aumento ni una disminución en el límite de unidades permitidas.

La implementación de estas tres recomendaciones en conjunto permite una reducción potencial de hasta el 10% en el número de unidades que el *marketplace* almacenaría en sus centros de distribución. Esto implica pasar de 1,350,800 unidades de SKUs nuevos de notebooks y celulares que podrían ingresar bajo la política actual a un total de 1,215,700 unidades.

Sin embargo, cabe destacar que durante el desarrollo del modelo se identificaron algunas limitaciones que podrían afectar su efectividad en la gestión del inventario del *marketplace* en cuestión. Una de las limitaciones clave es la dificultad para aplicar el modelo a productos poco estandarizados, debido a la diversidad en sus atributos y la falta de datos históricos consistentes. Además, el modelo no es capaz de anticipar eventos disruptivos que no están presentes en los datos de entrenamiento, lo que podría llevar a decisiones inapropiadas en la gestión del inventario. Por último, la naturaleza estática del modelo significa que no captura las fluctuaciones en las tendencias de demanda y la rápida obsolescencia de ciertos productos.

En resumen, el desarrollo la presente tesis representa un paso significativo hacia la optimización de la gestión de inventario en los centros de *fulfillment* del *marketplace* más grande de Latinoamérica. Al basar las decisiones sobre políticas de inventario en datos y análisis precisos, se abre la puerta a una operación más eficiente, una mejor alineación con las demandas del mercado y, en última instancia, un crecimiento sostenido para el negocio. Aunque se reconoce que existen limitaciones y desafíos por superar, este trabajo sienta las bases para futuras investigaciones y mejoras en la gestión del inventario, consolidando así el papel del *marketplace* como líder en el panorama del comercio electrónico en América Latina.

8 Referencias

Beneficios de la membresía Prime. (s.f.). Amazon. https://www.amazon.com/b/node=23945845011

Bergstra, J. y Bengio, Y. (2012). *Random search for hyper-parameter optimization.* Journal of Machine Learning Research.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.

Coppola, D. (2022). *E-commerce as Share of Total Retail Sales Worldwide 2015-2021 with Forecasts to 2026.* Statista.

Demand Forecasting: 10 key complexity drivers. (s.f.). Quantics. https://quantics.io/resource/demand-forecasting-10-key-complexity-drivers

Dieckmann, J. (2023). *Ensemble Learning: Bagging and Boosting*. Towards Data Science. https://towardsdatascience.com/ensemble-learning-bagging-and-boosting-23f9336d3cb0

Dietterich, T. G. (1995). *Overfitting and under-computing in machine learning.* ACM Computing Surveys.

Ecker, T., Hans, M., Neuhaus, F., y Spielvogel, J. (2020). *Same-day delivery: Ready for takeoff.* McKinsey. https://www.mckinsey.com/industries/retail/our-insights/same-day-delivery-ready-for-takeoff

El impacto del telégrafo en el comercio y la economía global. (s.f.). Historioteca. https://historioteca.com/el-impacto-del-telegrafo-en-el-comercio-y-la-economia-global

Estadísticas de Amazon Prime 2023: ¿Cuántas personas usan Amazon Prime? (s.f.). Pctg. https://pctg.net/estadisticas-de-amazon-prime-2023-cuantas-personas-usan-amazon-prime/

Fatemi, Z., Huynh, M., Zheleva, E., Syed, Z., & Di, X. (2023). *Mitigating Cold-start Forecasting using Cold Causal Demand Forecasting Model.* [Archivo PDF]. https://arxiv.org/pdf/2306.09261

FedEx Annual Report (2005). [Archivo PDF] https://s21.q4cdn.com/665674268/files/doc_financials/annual/2005/2005annualreport.pdf

Fiedler, L., Hazan, E., Ruwadi, B., y Ungerman, K. (2020). *Retail reimagined: The new era for customer experience.* [Archivo PDF]. https://www.mckinsey.com/~/media/McKinsey/Business%20Functions/Marketing%20and%20Sales/Periscope/Insights/Survey%20Reports/Reinventing%20Retail/Periscopes%20Retail%20Reimagined%20Report%202020-August-2020.pdf

Found, N., Lim, R., Holmes, J., y Duoanla, J. (2022). *Ecommerce Delivery Benchmark Report 2022. Welcome to the age of ecommerce*. [Archivo PDF]. https://info.metapack.com/rs/700-ZMT-762/images/Ecommerce%20Delivery%20Benchmark%20Report%202022%20%282%29.pdf

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian Data Analysis.* CRC Press.

Global E-Commerce Market Size & Share Analysis - Growth Trends & Forecasts (2023 - 2028). (s.f.). MordorIntelligence. https://www.mordorintelligence.com/industry-reports/global-ecommerce-market

Gitlan, D. (2024). *SSL History: From Inception to Evolutionary Triumph*. SSLDragon. https://www.ssldragon.com/blog/history-of-ssl/

Gonzalez, L. (2019). *Curvas ROC y área bajo la curva (AUC)*. Aprendeia. https://aprendeia.com/curvas-roc-y-area-bajo-la-curva-auc-machine-learning/

Goodfellow, I., Bengio, Y., y Courville, A. (2016). *Deep Learning.* MIT Press.

Greene, W. H. (2018). *Econometric Analysis*. Pearson.

Han, J., Kamber, M., y Pei, J. (2011). *Data mining: Concepts and techniques (3rd ed.).* Morgan Kaufmann.

Hastie, T., Tibshirani, R., y Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* Springer Science & Business Media.

Historia del Dinero. (s.f.). En *Wikipedia*. Recuperado el 01 de junio de 2024 de https://es.wikipedia.org/wiki/Historia_del_dinero

How Covid-19 triggered a Latin American e-commerce boom. (2021). OxfordBusinessGroup. https://oxfordbusinessgroup.com/articles-interviews/how-covid-19-triggered-a-latin-american-e-commerce-boom-2

Hyndman, R. J., y Koehler, A. B. (2006). *Another look at measures of forecast accuracy.* ScienceDirect. https://www.sciencedirect.com/science/article/abs/pii/S0169207006000239

Joshua, J. M. (2022). *El comercio en la antigua Mesopotamia.* WorldHistory. https://www.worldhistory.org/trans/es/2-2114/el-comercio-en-la-antigua-mesopotamia/

La historia de Amazon, el e-commerce gigante que revolucionó la industria. (2023). TheLogisticsWorld. https://thelogisticsworld.com/logistica-comercio-electronico/la-historia-de-amazon-el-e-commerce-gigante-que-revoluciono-la-industria

Lee H., Kim S.G., Park H. W., Kang P. (2014). *Pre-launch new product demand forecasting using the Bass model: A statistical and machine learning-based approach.* Technological Forecasting & Social Change.

Lopienski, K. (2023). *Cross-docking explained: Here's what you need to know.* Shipbob. https://www.shipbob.com/blog/cross-docking/#h-what-is-cross-docking

Machine Learning: AI's Role in the Future of eCommerce. (2021). VARStreet. https://blog.varstreetinc.com/machine-learning-ais-role-in-the-future-of-ecommerce

Marko. (2021). *E-commerce in Post-Pandemic World.* Younify. https://www.younify.eu/e-commerce-in-post-pandemic-world/

Measuring the value of e-commerce. (2023). UNCTAD. [Archivo PDF]. https://unctad.org/system/files/official-document/dtlecde2023d3_en.pdf

Mehta, D. (2023). *Fulfillment by Amazon: How improving delivery fueled independent seller growth and success*. AboutAmazon. https://www.aboutamazon.com/news/small-business/fulfillment-by-amazon-how-improving-delivery-fueled-independent-seller-growth-and-success

Mercaderes y Ferias en la Edad Media: Rutas Comerciales. (s.f.). HistoriaYBiografias. https://historiaybiografias.com/ferias/

Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective.* MIT Press.

Nair, B., Gupta, G., Dadarkar, M., y Nandamuri, B. (2022). *Generate cold start forecasts for products with no historical data using Amazon Forecast, now up to 45% more accurate*. Amazon. https://aws.amazon.com/es/blogs/machine-learning/generate-cold-start-forecasts-for-products-with-no-historical-data-using-amazon-forecast-now-up-to-45-more-accurate/

Naveira, A. (2020*). Historia de eBay: nacimiento y evolución de uno de los mayores marketplaces del mundo*. Marketing4Ecommerce. https://marketing4ecommerce.mx/historia-ebay/

OneVsRestClassifier. (s.f.). ScikitLearn. https://scikit-learn.org/stable/modules/multiclass.html#onevsrestclassifier

Oxford Learner's Dictionaries. (s.f.). Marketplace. En *Oxfords Advanced Learner's Dictionary*. Recuperado el 1 de junio de 2024, de https://www.oxfordlearnersdictionaries.com/definition/english/marketplace?q=marketplace

Perrin, A., y Duggan, M. (2015). *American's Internet Access: 2000-2015.* PewResearch. https://www.pewresearch.org/internet/2015/06/26/americans-internet-access-2000-2015/

Pop, A. (2022). *Una guía sobre los diversos tipos de marketplaces.* Vtex. https://vtex.com/latam/blog/estrategia-latam/tipos-de-marketplaces-guia/

Raschka, S., y Mirjalili, V. (2019). *Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow 2*. Packt Publishing.

Rocca, J. (2019). *Ensemble methods: Bagging, Boosting, and Stacking. Understanding the key concepts of ensemble learning.* TowardsDataScience. https://towardsdatascience.com/ensemble-methods-bagging-boosting-and-stacking-c9214a10a205

Rodriguez, J. (2023). *Qué es un marketplace: ejemplos, tipos y cómo funcionan.* Hubspot. https://blog.hubspot.es/sales/que-es-marketplace

Role of Data Analytics in Logistics. (2023). LogisticsPeek. https://logisticspeek.com/blog/role-of-data-analytics-in-logistics/

Schafer, J. B., Frankowski, D., Herlocker, J., y Sen, S. (2007). *Collaborative Filtering Recommender Systems.* Springer.

Seyedan, M., y Mafakheri, F. (2020). *Predictive big data analytics for supply chain demand forecasting: methods, applications, and research opportunities*. JournalOfBigData. https://journalofbigdata.springeropen.com/articles/10.1186/s40537-020-00329-2

Shaw, N., Eschenbrenner, B., y Baier, D. (2022). Online shopping continuance after COVID-19: A comparison of Canada, Germany, and the United States. NCBI. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9379614/

Stanley, L. (2022). *The Complete History of Ecommerce*. Nexcess. https://www.nexcess.net/blog/history-ecommerce/

The Evolution of Digital Payments and E-Commerce. (2023). FinanceMagnates. https://www.financemagnates.com/fintech/payments/the-evolution-of-digital-payments-and-e-commerce/

Thonemann, U., y Bradley, J. (2002). *The Effect of Product Variety on Supply-Chain Performance.* European Journal of Operational Research.

Tian, Y. y Stewart, C. (2007). *History of e-commerce. In Electronic Commerce: concepts, methodologies, tools, and applications.* IGI Global.

We are Mercado Libre. (s.f.). [Archivo PDF]. https://api.mziq.com/mzfilemanager/v2/d/098a2d95-0ea8-4ed5-a340-d9ef6a2b0053/bed27527-5828-d2c9-9e35-1e5e92d418ee?origin=2

What Is An Online Marketplace? (s.f.). Stockarea. https://stockarea.io/blogs/online-marketplace/

What is Dropshipping? Benefits, Challenges, and Getting Started. (2020). Adobe. https://business.adobe.com/blog/basics/dropshipping

9 Apéndice

9.1 Listado de Variables en cada *Dataset*

***Dataset* de ítems:** para cada combinación de *ítem_id* y *variation_id*, se disponibiliza la siguiente información actual:

- País.
- ID del vendedor.
- ID de tienda oficial.
- Fecha de inicio de la publicación.
- Título de la publicación.
- Tipo de publicación (gratuita, clásica, *premium*).
- Comisión por venta.
- Condición (usado, nuevo).
- Estatus de la publicación *(active, paused, under review, deleted, inactive, closed, payment required, never paid, suspended by user, not yet active*).
- ID de la categoría.
- ID del dominio.
- *Flag* publicación catálogo.
- ID publicación catálogo.
- *Flag* promoción.
- Garantía.
- Precio.
- Moneda.
- Dimensiones.
- Tipo de envío disponible *(Fulfillment, Cross Docking, Drop off, Self Service, Not Specified).*
- *Flag* de *pick-up in store.*
- ID de dirección del vendedor.
- Código postal vendedor.
- Ciudad del vendedor.
- Estado del vendedor.
- Cantidad de fotos.

- Cantidad de variaciones del ítem.
- Cantidad de atributos cargados (características del producto).
- ID del atributo.
- Nombre del atributo.
- Valor del atributo.

***Dataset* de actividad del ítem:** contiene la misma información que el *dataset* de ítems (a) para cada combinación de *item_id* y *variation_id*, pero guardando un registro por cada modificación que tuvo la publicación:

- Fecha/hora de actualización.
- País.
- ID del vendedor.
- ID de tienda oficial.
- Fecha de inicio de la publicación.
- Título de la publicación.
- Tipo de publicación *(free, bronce, silver, gold special, gold premium, gold pro, gold*).
- Comisión por venta.
- Condición *(used, new, not specified).*
- Estatus de la publicación *(active, paused, under review, deleted, inactive, closed, payment required, never paid, suspended by user, not yet active).*
- ID de la categoría.
- ID del dominio.
- *Flag* publicación catálogo.
- ID publicación catálogo.
- *Flag* promoción.
- Garantía.
- Precio.
- Moneda.
- Dimensiones.
- Tipo de envío disponible *(Fulfillment, Cross Docking, Drop off, Self Service, Not Specified).*
- *Flag* de *pick-up in store.*

- ID de dirección del vendedor.
- Código postal vendedor.
- Ciudad del vendedor.
- Estado del vendedor.
- Cantidad de fotos.
- Cantidad de variaciones del ítem.
- Cantidad de atributos cargados (características del producto).
- ID del atributo.
- Nombre del atributo.
- Valor del atributo.

***Dataset* de reputación del vendedor:** para cada combinación de *cust_id* y *photo_id*, se disponibiliza la siguiente información:

- País.
- Nivel de reputación actual *(red, orange, yellow, light_green, green, green silver, green gold, green platinum).*
- Puntuaciones positivas por órdenes concretadas.
- Puntuaciones neutrales por órdenes concretadas.
- Puntuaciones negativas por órdenes concretadas.
- Puntuaciones neutrales por órdenes no concretadas.
- Puntuaciones negativas por órdenes no concretadas.
- Nivel de reclamos *(red, orange, yellow, light green).*
- Nivel de cancelaciones *(red, orange, yellow, light green).*
- Nivel de demoras en despacho *(red, orange, yellow, light green).*

***Dataset* de reputación del ítem:** para cada combinación de *item_id* y *photo_id*, se disponibiliza la siguiente información:

- País.
- ID del vendedor.
- Nivel de reputación actual *(red, yellow, green).*
- Cantidad de reclamos.
- Nivel de reclamos *(red, yellow, green, newbie).*

- Cantidad de cancelaciones.
- Nivel de cancelaciones *(red, yellow, green, newbie).*
- Salud de reputación *(newbie, unhealthy, warning, healthy).*

***Dataset* de ventas:** para cada *order_id*, se disponibliza la siguiente información:

- Fecha / hora de compra.
- País.
- ID Vendedor.
- ID Comprador.
- Ítem.
- Variación.
- Status (*canceled, payment in process, confirmed, cancelled, payment required, invalid, partially paid, closed, partially refunded, pending cancel, paid).*
- Status del pago.
- Método de pago (tarjeta de crédito, débito, atm, dinero en cuenta).
- Cuotas.
- Monto total de la orden ($).
- Monto total pagado ($).
- Moneda.
- Cantidad de ítems en la orden.
- Precio por unidad del ítem.
- Descuento por campaña (%).
- ID campaña de descuento.
- ID cupón.
- Monto cupón.
- Método.
- Comisión.
- Tipo de publicación.
- Elegible para *Buy Box.*
- Garantía.
- Envío seleccionado *(Fulfillment, Cross Docking, Drop off, Self Service, Not Specified).*

- *Flag pickup.*
- ID del envío.
- Cargos por envío.
- *Flag free shipping.*

9.2 Variables Incluidas en *Dataset* de Entrenamiento

***Features* del vendedor:** para cada CUS_CUST_ID:

- 'PAIS' [object].
- 'FLAG_TIENDA_OFICIAL' [int64].
- 'SELLER_REP_CURRENT_LEVEL' [object].
- 'VENTAS_HISTORICAS_VENDEDOR' [int64].
- 'ITE_ITEM_SELLER_ADDRESS_ID' [int64].
- 'CODIGO_POSTAL' [float64].
- 'CITY' [object].
- 'STATE' [object].

***Features* del ítem (publicación):** para cada ITE_ITEM_ID:

- 'ITE_ITEM_TITLE' [object].
- 'CATEG_ID' [int64].
- 'CATEG_NAME' [object].
- 'FLAG_CATALOGO' [int64].
- 'PRODUCT_ID' [float64].
- 'DOMAIN' [object].
- 'LISTING_TYPE' [object].
- 'FLAG_PROMOTION_TODAY' [int64].
- 'BUYING_MODE' [object].
- 'PRICE_USD' [float64].
- 'PRICE_LC' [float64].
- 'SIT_CURRENCY_ID' [object].
- 'ITE_ITEM_CONDITION' [object].
- 'FLAG_PICKUP_TIENDA' [int64].

- ‘LOGISTIC_TYPE’ [object].
- ‘FLAG_ENVIO_GRATIS’ [int64].
- ‘FLAG_GARANTIA’ [int64].
- ‘FLAG_VIDEO’ [int64].
- ‘PICTURE_QTY’ [int64].
- ‘CANT_VARIACIONES’ [int64].
- ‘ATTRIBUTE_QTY’ [int64].
- ‘ITE_REP_CURRENT_LEVEL’ [object].
- ‘ITE_REP_CLAIMS_LEVEL’ [object].
- ‘ITE_REP_CLAIMS_HEALTH’ [object].
- ‘ITE_REP_CANCEL_LEVEL’ [object].
- ‘ITE_REP_CANCEL_HEALTH’ [object].
- ‘ITE_REP_HEALTH’ [object].

***Features* de la variante:** para cada VARIATION_ID:

- ‘AVAILABLE_QTY’ [int64].
- ‘BRAND’ [object].
- ‘MODEL’ [object].
- ‘DETAILS’ [object].
- ‘LINE’ [object].
- ‘EAN’ [object].
- ‘DISPLAY_TYPE’ [object].
- ‘PROCESSOR_MODEL’ [object].
- ‘COLOR’ [object].
- ‘DISPLAY_SIZE’ [object].
- ‘DISPLAY_RESOLUTION’ [object].
- ‘PROCESSOR_SPEED’ [object].
- ‘WIDTH_cm’ [float64].
- ‘HEIGHT_cm’ [float64].
- ‘DEPTH_cm’ [float64].
- ‘VOLUMEN_cm3’ [float64].

- ‘WEIGHT_grs’ [float64].

9.3 Correlación entre *Features*

Entre las correlaciones positivas entre *features* relevantes cabe destacar:

- La correlación perfecta entre la cantidad de minutos que la publicación estuvo activa (‘*minutos_activa_N6W*’) y la cantidad de días que la publicación estuvo activa (‘*dias_activa_N6W*’). Sin embargo, ambas variables explican la misma información, por lo que puede considerarse redundante.
- La correlación entre la cantidad de atributos de la publicación (‘*cant_atributos*’) y el hecho de que la publicación esté catalogada (*‘flag_catalogo’*). Esto indica que cuanta más información se cargue en la publicación, mayor es la probabilidad de que dichos datos provengan del equipo de catálogo y no del vendedor (corr: +0,31).
- La correlación entre la posibilidad de retirar el producto por la tienda (*‘flag_ pickup_tienda’*) y el código postal del vendedor (*‘codigo_postal’*) sugiere que permitirle al comprador evitar el costo de envío o adquirir su compra en el momento, está positivamente relacionada con la ubicación geográfica del vendedor (corr: +0,29).

Entre las correlaciones negativas entre *features* relevantes cabe destacar:

- La correlación negativa entre el flag de información cargada por el equipo de catálogo (*‘flag_catalogo’*) y la cantidad de variantes de la publicación (*‘cant_variaciones’*). Esto sugiere que a medida que aumenta la cantidad de variantes de una publicación, es menos probable que la información haya sido proporcionada por el equipo de catálogo (corr: -0,84). Esto tiene sentido ya que las publicaciones muy personalizadas no suelen ser catalogadas.
- La correlación negativa entre la cantidad de variantes de la publicación (*‘cant_variaciones’*) y la cantidad de atributos cargados en la misma (*‘cant_atributos’*), ya que los vendedores suelen cargar menos especificaciones que las que carga el equipo de catálogo (corr: -0,23).
- La correlación negativa entre *‘flag_ pickup_tienda’* y el *‘flag_envio_gratis’*, lo que indica que es más probable que una publicación que permite el retiro en tienda no ofrezca envíos gratuitos (corr: -0,21). Esto podría explicarse en productos donde el *ticket* promedio es más bien bajo, y por lo tanto no resulta rentable para el vendedor ofrecer envíos sin costo.
- La correlación negativa entre *‘flag_ pickup_tienda’* y *‘flag_tienda_oficial’* (corr: -0,18), sugiriendo que es menos probable que los vendedores más grandes e importantes de la plataforma habiliten

la posibilidad de retirar por tienda, sino que optan por ofrecer una mejor experiencia de compra con envío sin cargo.

9.4 Relación entre *Features* y Variable *Target*

Estado del vendedor

Las Figura 9.1 muestran los estados de cada país que concentran el 95% de la demanda de los SKUs nuevos. En Brasil, la demanda se concentra en el estado de São Paulo, con un 80% del total nacional. Paraná, por otro lado, tiene una participación mucho menor, con solo el 7% de la demanda. En México, la demanda está más distribuida entre los estados debido a su estructura federal. El Distrito Federal lidera con el 56% de las ventas, seguido por el Estado de México con el 20%. Los estados de Morelos y Jalisco también son significativos, cada uno con el 6% de las ventas. Estos estados juntos explican el 88% de la demanda, mientras que otros estados contribuyen en menor medida.

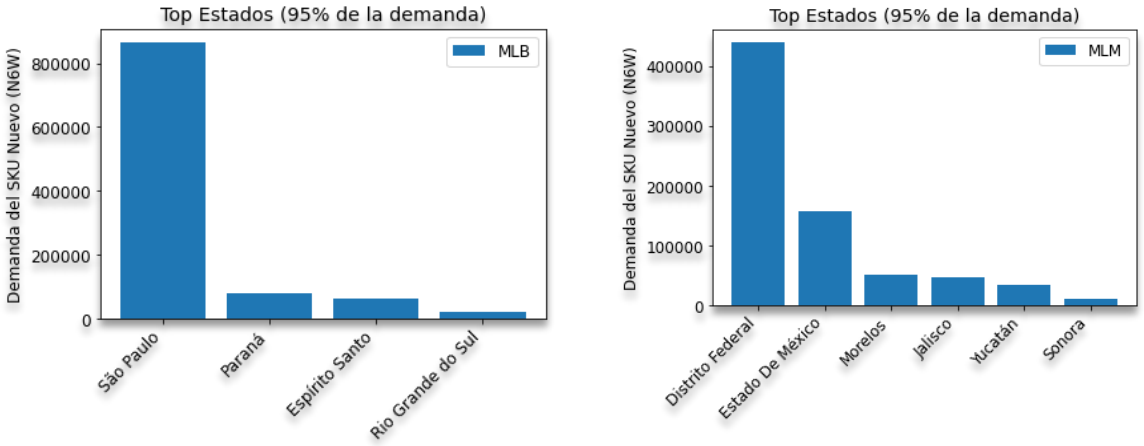


Figura 9.1 Izquierda: Top estados Brasil (95% de la demanda). Derecha: Top estados México (95% de la demanda)

Esta variable no es relevante para el modelo, ya que cuando las publicaciones ofrecen *fulfillment* la plataforma se encarga de redistribuir el *stock* por todo el país. Por lo tanto, las variables relacionadas con la ubicación del vendedor pierden importancia para estimar las ventas de un producto nuevo.

Ciudad del vendedor

Se analiza la relación entre el campo *'ciudad'* y la variable objetivo. Las Figura 9.2 muestran las ciudades de Brasil y México, que concentran el 80% de la demanda de los SKUs nuevos.

En Brasil, la demanda de SKUs nuevos se concentra notablemente en la ciudad de São Paulo, que acapara el 42% del total nacional. La ciudad de Cajamar sigue con el 17% de la demanda. Las demás ciudades aportan menos del 5% cada una.

En México, la distribución de la demanda de productos nuevos es más equitativa. Miguel Hidalgo lidera con el 23% de las unidades vendidas, seguida por Tepotzotlán y Cuauhtémoc, ambas con el 13%. Otras ciudades contribuyen con menos del 8% de las ventas cada una. Estas ciudades representan conjuntamente el 80% de la demanda en México, aunque otras ciudades también registran ventas, aunque en menor proporción.

Similar al análisis del estado del vendedor, la variable *'ciudad'* no parece ser de gran utilidad para estimar la demanda de un producto nuevo que ingresará a *fulfillment*, ya que el *marketplace* redistribuye el *stock* a lo largo del país.

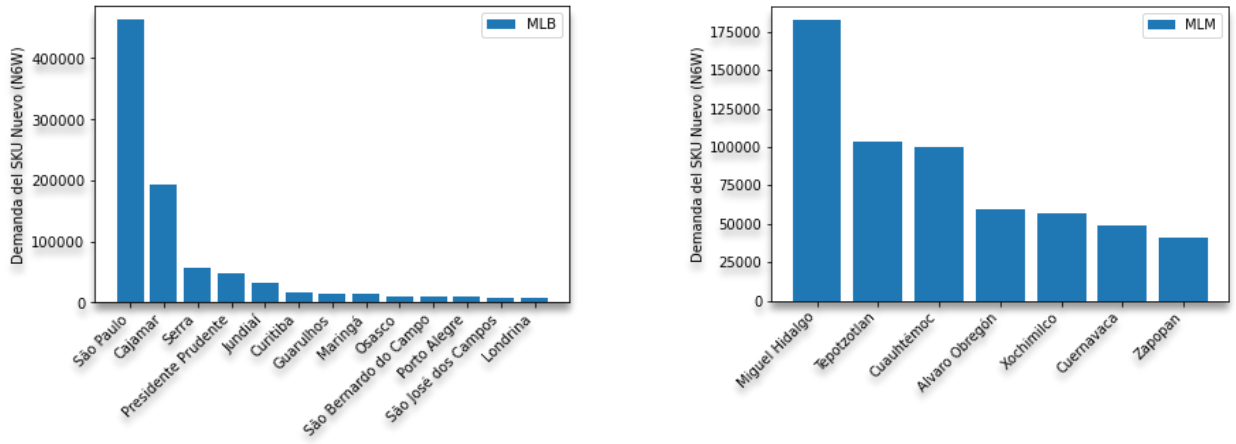


Figura 9.2 Izquierda: Top ciudades de Brasil (80% de la demanda), Derecha: Top ciudades de México (80% de la demanda)

Código postal del vendedor

Las Figuras 9.3 y 9.4 muestran los códigos postales con la mayor demanda de SKUs nuevos en Brasil y México, respectivamente, junto con la ciudad y el estado correspondientes. Es importante destacar que el código postal mencionado se refiere al vendedor (origen de la compra), no al comprador (destino de la compra).

En ambos países, hay una concentración significativa de vendedores en los estados más grandes, como São Paulo y el Distrito Federal. Esto sugiere que los vendedores en las principales ciudades pueden ofrecer tiempos de entrega más cortos, lo que podría aumentar la demanda de sus productos.

En Brasil, los códigos postales con la mayor demanda se encuentran en São Paulo y Cajamar, destacándose los códigos 4571900, 2185031 y 5035901 en São Paulo, y 7750020 en Cajamar.

En México, los códigos postales con mayor demanda se encuentran en la delegación Miguel Hidalgo (11230), Tepotzotlán (54616) y Álvaro Obregón (1120).

Similar a las variables *'país'* y *'ciudad'*, el código postal no es relevante para el modelo, ya que al ingresar a *fulfillment*, la ubicación del vendedor deja de ser importante para estimar la demanda de un producto nuevo.

País	Código postal	Ciudad	Estado	% demanda N6W	% Acumulado
MLB	7750020	Cajamar	São Paulo	17,4%	17,4%
MLB	4571900	São Paulo	São Paulo	9,4%	26,8%
MLB	2185031	São Paulo	São Paulo	4,3%	31,1%
MLB	5035901	São Paulo	São Paulo	3,3%	34,4%
MLB	29161384	Serra	Espírito Santo	3,0%	37,4%
MLB	3116000	São Paulo	São Paulo	2,7%	40,1%
MLB	1422001	São Paulo	São Paulo	2,3%	42,4%
MLB	13213086	Jundiaí	São Paulo	1,7%	44,2%
MLB	1031001	São Paulo	São Paulo	1,7%	45,9%
MLB	29173795	Serra	Espírito Santo	1,2%	47,1%
MLB	13213090	Jundiaí	São Paulo	1,0%	48,1%
MLB	12232360	São José dos Campos	São Paulo	0,7%	48,8%
MLB	4220000	São Paulo	São Paulo	0,7%	49,5%

Figura 9.3 Top código postal de Brasil (50% de la demanda)

País	Código postal	Ciudad	Estado	% demanda N6W	% Acumulado
MLM	11230	Miguel Hidalgo	Distrito Federal	15,8%	15,8%
MLM	54616	Tepotzotlán	Estado De México	8,9%	24,7%
MLM	1120	Álvaro Obregón	Distrito Federal	7,5%	32,3%
MLM	6600	Cuauhtémoc	Distrito Federal	6,9%	39,2%
MLM	62374	Cuernavaca	Morelos	6,3%	45,4%
MLM	16000	Xochimilco	Distrito Federal	4,4%	49,9%
MLM	11230	Miguel Hidalgo	Distrito Federal	15,8%	15,8%
MLM	54616	Tepotzotlán	Estado De México	8,9%	24,7%

Figura 9.4 Top código postal de México (25% de la demanda)

Flag catálogo

En el *marketplace* analizado, una publicación se considera parte del catálogo si está asociada a un producto existente en la plataforma. El catálogo permite que varias publicaciones del mismo producto compitan por destacarse como la mejor oferta, considerando factores como el precio, la disponibilidad de cuotas sin interés y la reputación del vendedor. Las publicaciones de tipo catálogo suelen tener fotos de alta calidad y descripciones detalladas, mejorando la experiencia de compra.

En el *dataset* analizado, el 43% de las observaciones son publicaciones del catálogo. Sin embargo, la Figura 9.5 muestra que no hay una relación clara entre el *flag* de catálogo y la demanda del SKU nuevo en las primeras seis semanas. Esto indica que, aunque el catálogo ayuda a competir en los resultados de búsqueda, otros factores como la calidad del producto y la estrategia de marketing del vendedor son determinantes en las ventas.

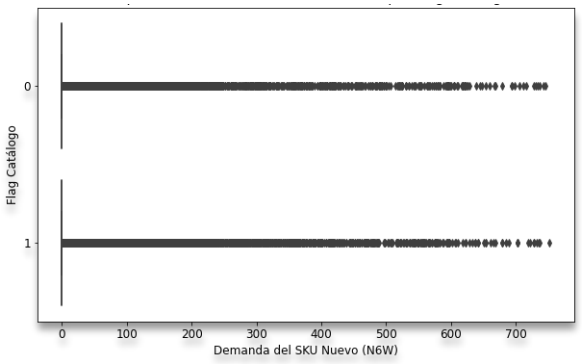


Figura 9.5 Boxplot de demanda del SKU nuevo según flag catálogo

Flag *pickup* en tienda

Para mejorar la experiencia de compra, la plataforma permite a los vendedores indicar en sus publicaciones si los productos pueden ser retirados en tienda. Esta opción brinda a los compradores la flexibilidad de elegir el lugar de retiro de su pedido y facilita la preparación por parte del vendedor.

En el análisis de datos, se encontró que el 31% de las publicaciones tienen activada la opción de retiro en tienda. Contrario a lo que se podría esperar, las publicaciones que ofrecen esta opción suelen tener una demanda menor en comparación con aquellas que no la ofrecen, como muestra la Figura 9.6. Esto podría deberse a que los compradores prefieren la comodidad del envío a domicilio o a que la opción de retiro en tienda no es un factor decisivo para la mayoría de los consumidores.

A priori, esta variable no parece tener información relevante para el modelo en cuestión, ya que la disponibilidad de retiro en tienda no parece influir significativamente en la demanda del SKU nuevo.

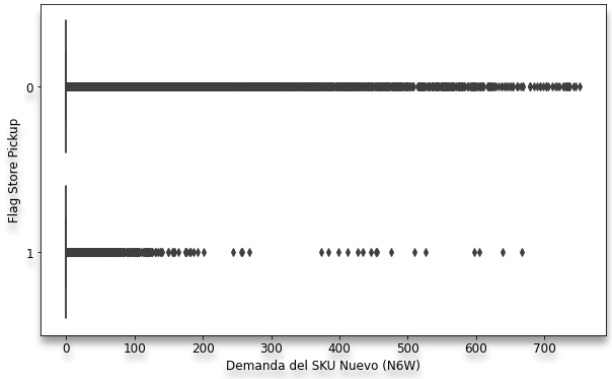


Figura 9.6 Boxplot de demanda del SKU nuevo según flag pickup en tienda

Flag envío gratis

Los envíos gratuitos son un beneficio muy valorado por los consumidores. La plataforma ofrece esta opción para las publicaciones de productos nuevos con un valor superior a 79 Reales en Brasil y 299 pesos en México. Aunque el costo del envío gratis para el vendedor depende del peso y destino del paquete, la plataforma brinda descuentos según la reputación del vendedor, que pueden cubrir hasta el 50% del costo.

El análisis del *dataset* revela que el 91% de las observaciones ofrecen envío gratis, como se muestra en la Figura 9.7 a la izquierda. Además, se observa a la derecha que los SKUs nuevos con envío gratis acaparan más del 99% de la demanda total.

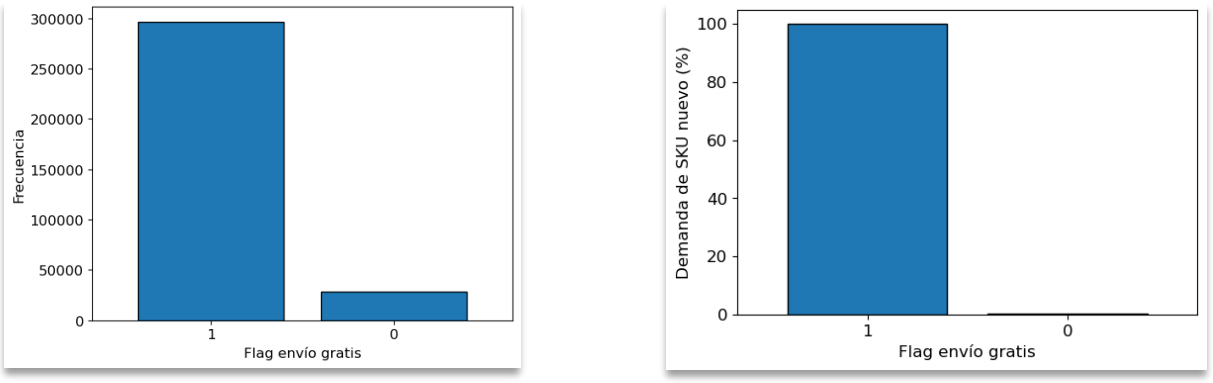


Figura 9.7 Izquierda: Distribución de observaciones según flag de envío gratis. Derecha: Distribución de demanda de SKUs nuevos según flag de envío gratis

Sin embargo, para el modelo en cuestión, esta variable no resulta de mucho interés, ya que, en los dominios de *notebooks* y celulares, el *ticket* promedio supera el umbral mínimo para ofrecer envío gratis de forma obligatoria por la plataforma. Los casos donde el valor no supera este umbral son atípicos en la muestra.

Flag garantía

En la plataforma, los vendedores pueden ofrecer garantías a los consumidores para generar confianza en las compras, ya sean de fábrica o proporcionadas por el vendedor.

La Figura 9.8 a la izquierda muestra que el 98% de las publicaciones ofrecen algún tipo de garantía, mientras que solo el 2% no. Además, la Figura 9.8 a la derecha indica que las publicaciones con garantía presentan una demanda significativamente mayor en las primeras seis semanas.

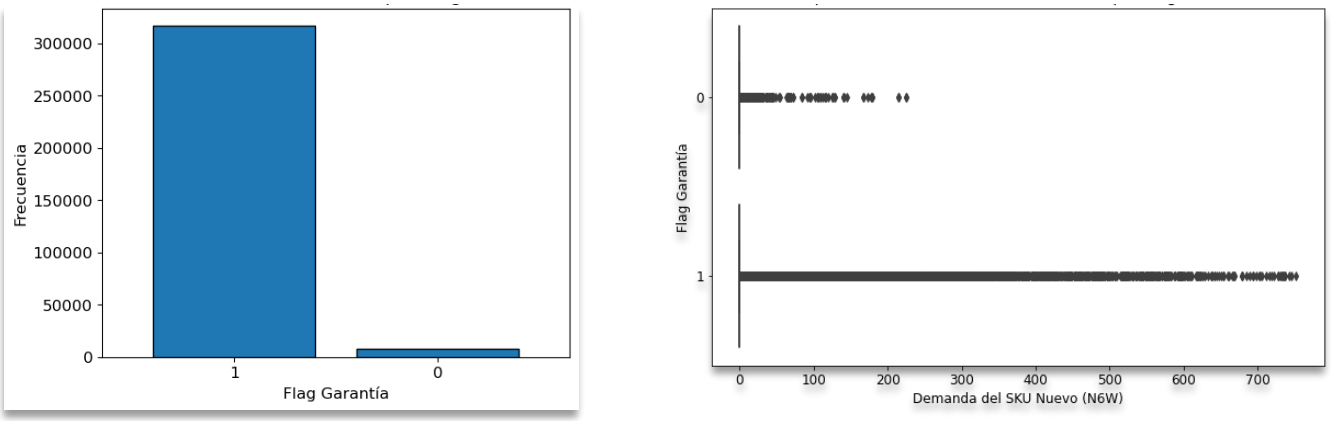


Figura 9.8 Izquierda: Distribución de observaciones según flag de garantía. Derecha: Boxplot de demanda del SKU nuevo según flag de garantía

Sin embargo, en los dominios de *notebooks* y celulares, la regulación exige que los fabricantes brinden algún tipo de garantía en caso de fallas o desperfectos del producto. Por lo tanto, todos los SKUs nuevos que ingresen a *fulfillment* deberían cumplir con este requisito. Dado que la garantía es obligatoria en estos casos, no resulta útil como variable para el modelo al momento de estimar la demanda esperada del producto.

Flag video

La plataforma permite a los vendedores subir videos que muestren el producto, con el objetivo de aumentar la confianza de los consumidores y mejorar la calidad de la publicación.

Sin embargo, la Figura 9.9 a la izquierda muestra que solo el 6% de las observaciones cuentan con un video cargado, indicando que no es una herramienta ampliamente utilizada por los vendedores.

Además, la Figura 9.9 a la derecha revela que, aunque los videos pueden ser útiles para el comprador, no se observan diferencias estadísticamente significativas en la demanda de un SKU nuevo según si tiene o no un video publicado en la publicación. Por este motivo, la variable no parecería aportar mucha información al momento de detectar patrones que determinen la demanda en este tipo de productos.

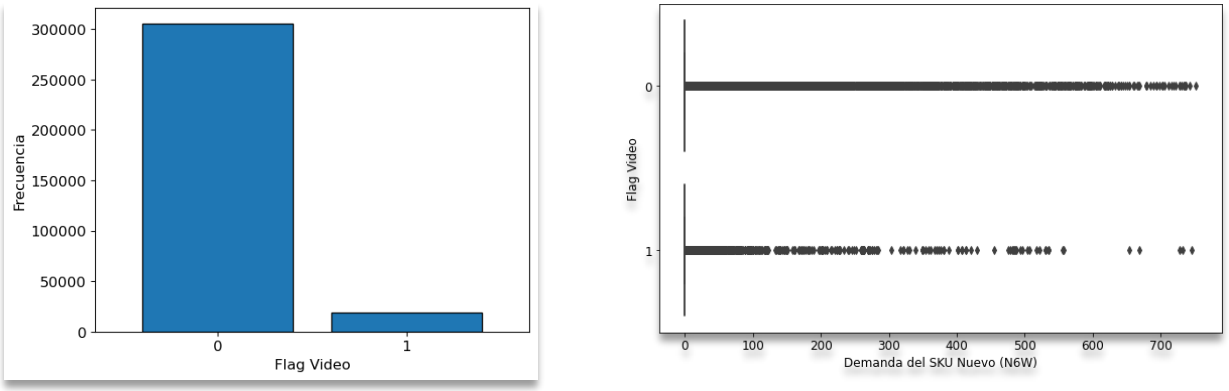


Figura 9.9 Izquierda: Distribución de observaciones según flag video. Derecha: Distribución de demanda del SKU nuevo según flag de video

Cantidad de fotos de la publicación

La Figura 9.10 izquierda muestra la distribución de observaciones según la cantidad de fotos en la publicación, con una media de 5 fotos. A la derecha, la demanda aumenta con la cantidad de fotos, alcanzando su punto máximo en publicaciones con 6 fotos. Posteriormente, la demanda disminuye progresivamente, y las publicaciones con más de 11 fotos prácticamente no concentran ventas.

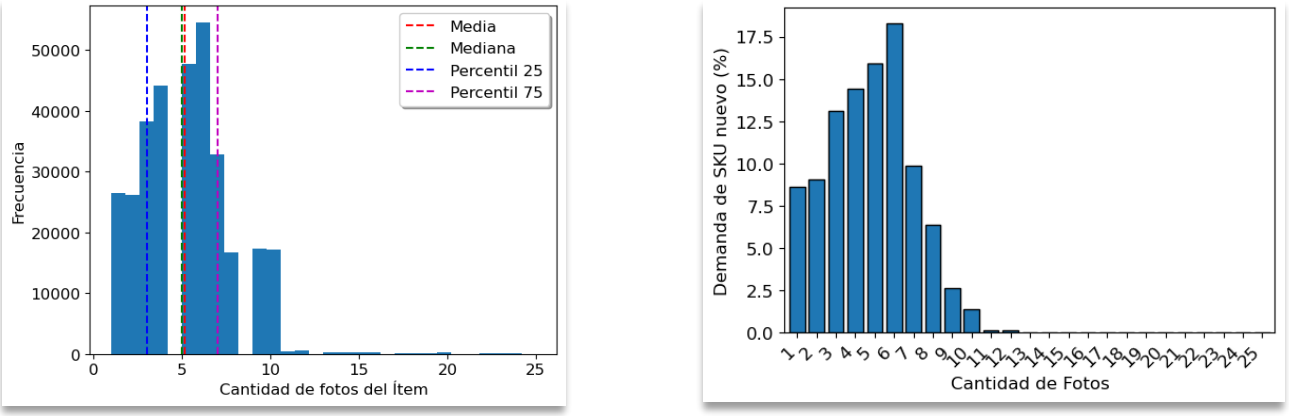


Figura 9.10 Izquierda: Distribución de observaciones según cantidad de fotos de la publicación. Derecha: Distribución de la demanda de SKUs nuevos según cantidad de fotos de la publicación

Sin embargo, en los modelos analizados, la cantidad de fotos no resulta ser una variable significativa para estimar la demanda de un SKU nuevo. Por lo tanto, es importante considerar la calidad de las fotos, y no solo la cantidad, al crear publicaciones atractivas para los compradores.

Cantidad de variantes de la publicación

En la plataforma, es común que las publicaciones presenten distintas variantes de un producto, especialmente en categorías como celulares y *notebooks*, donde suelen variar en color, cantidad de memoria, tamaño, entre otros.

La Figura 9.11 izquierda muestra la distribución de las observaciones del *dataset* en función de la cantidad de variantes de la publicación. Se observa que aproximadamente el 50% de las publicaciones no presentan ninguna variante, mientras que cerca del 48% presentan como máximo una. Esto indica que en esta categoría de productos no es común encontrar publicaciones con varias variantes.

La Figura 9.11 a la derecha muestra que los ítems con dos o más variantes tienen menos ventas en las primeras seis semanas en comparación con aquellos que no tienen ninguna variante o tienen como máximo una variante.

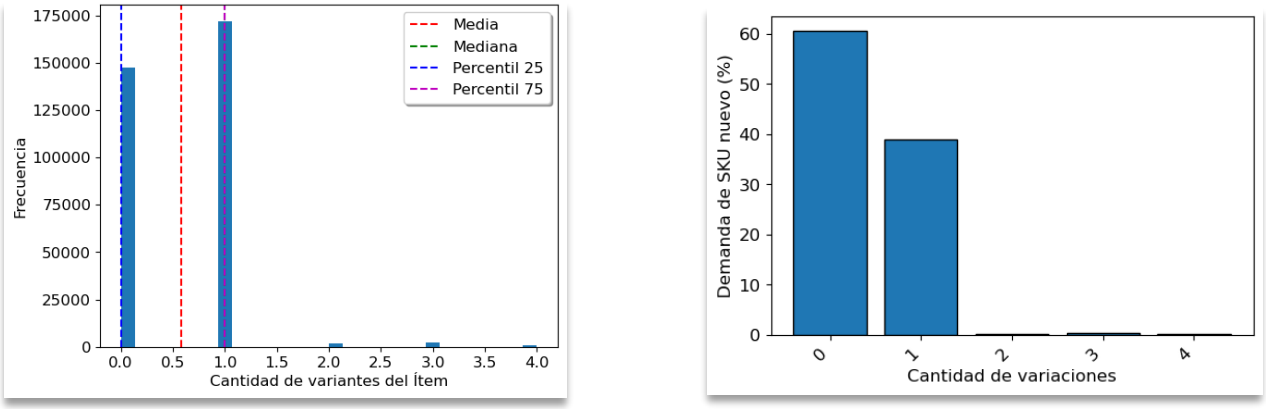


Figura 9.11 Izquierda: Distribución de observaciones según cantidad de variantes de la publicación. Derecha: Distribución de la demanda de SKUs nuevos según cantidad de variantes de la publicación

Dado que no se observan grandes diferencias en la demanda de productos con ninguna o una variante, y hay pocos casos donde este tipo de dominio presenta más de una variante, esta variable no parece ser de gran valor para estimar la demanda de un producto.

Mes de creación de la publicación

Al examinar la cantidad de observaciones por mes en la Figura 9.12, se puede afirmar que la muestra es bastante equilibrada, con una pequeña concentración de observaciones en los meses anteriores a diciembre de 2021 (aproximadamente un 12% de las observaciones en cada mes), en comparación con los meses posteriores (aproximadamente un 8% de las observaciones en cada mes).

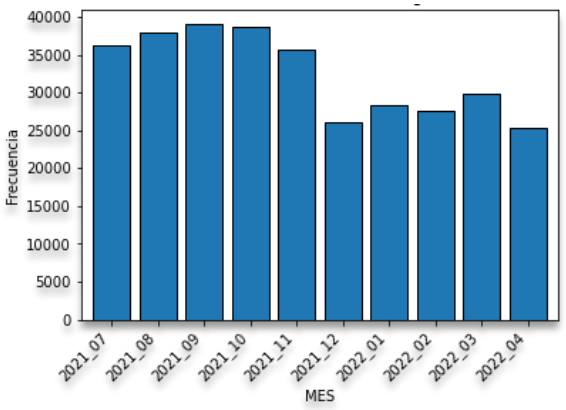


Figura 9.12 Distribución de observaciones según mes de publicación

Sin embargo, al estudiar la relación entre la demanda y el mes para identificar patrones de estacionalidad en la Figura 9.13, no se detecta una tendencia marcada en estas categorías. La media de la demanda por mes oscila entre 3,3 y 4,5 ventas en las primeras seis semanas.

Mes	Media	Mediana	Perc'25	Perc'50	Perc'75	Perc'90	Perc'95	Perc'99
Jul'21	3,9	0	0	0	0	4	14	92
Ago'21	3,8	0	0	0	0	4	14	91
Sep'21	3,3	0	0	0	0	4	11	75
Oct'21	3,3	0	0	0	0	4	12	71
Nov'21	3,4	0	0	0	0	3	10	83
Dic'21	3,9	0	0	0	0	4	14	96
Ene'22	3,6	0	0	0	0	4	14	81
Feb'22	3,5	0	0	0	0	4	14	75
Mar'23	4,3	0	0	0	0	6	17	99
Abr'23	4,5	0	0	0	0	4	17	109

Figura 9.13 Distribución de demanda de SKU nuevo en las primeras seis semanas según mes de publicación

Por este motivo la variable en cuestión aparentemente no tendría gran valor para el modelado del problema.

Volumen del SKU nuevo (cm3)

La volumetría del producto está directamente relacionada con la categoría; SKUs de menor volumen corresponden a celulares y los de mayor volumen a *notebooks*.

La Figura 9.14 izquierda muestra que el 88% de los SKUs nuevos tienen una volumetría inferior a los 230 cm^3, lo que equivale aproximadamente a un producto de 15 cm x 8 cm x 2 cm por lado (un *proxy* de las medidas de un celular). Esto sugiere que la mayoría de las observaciones del *dataset* son celulares, y no *notebooks* (o al menos del 80% que tienen información de las medidas). Sin embargo, es importante destacar que esta variable tiene aproximadamente un 1% de valores atípicos (superando los 2.348 cm^3), probablemente debido a errores de carga de la unidad de medida. También se observa que las medidas de dispersión (mediana y percentiles 25 y 75) oscilan entre los 100 cm^3 y los 115 cm^3.

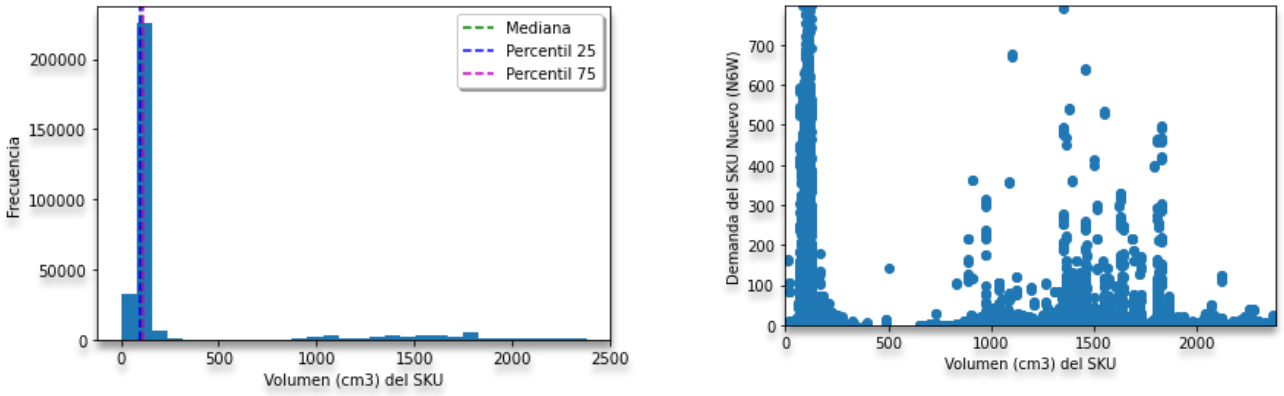


Figura 9.21 Izquierda: Distribución de observaciones según volumetría del SKU. Derecha: Distribución de la demanda según volumetría del SKU

La Figura 9.14 a la derecha analiza la relación entre las ventas de un SKU nuevo y su volumetría. Se observa una alta concentración de productos que alcanzan niveles de ventas más altos alrededor de los 100 cm^3.

Estos hallazgos son relevantes para el modelado del problema, ya que sugieren que la volumetría del SKU es un factor importante en la determinación de la demanda de un producto nuevo. Sin embargo, esta variable está estrechamente correlacionada con la categoría del producto, ya que indica el tipo de producto (celular o *notebook*). Por lo tanto, la volumetría por sí sola puede no aportar información adicional significativa para el modelo en cuestión.

Peso del SKU nuevo (grs)

La Figura 9.15 a la izquierda muestra la distribución de observaciones según el peso del SKU nuevo. Se observa que el 88% de los productos tienen un peso inferior a los 408 gramos, sugiriendo que la mayoría de la base de datos corresponde a celulares y no a notebooks. Las demás observaciones alcanzan hasta un máximo de 2.300 gramos, descartando el 1% de los valores atípicos.

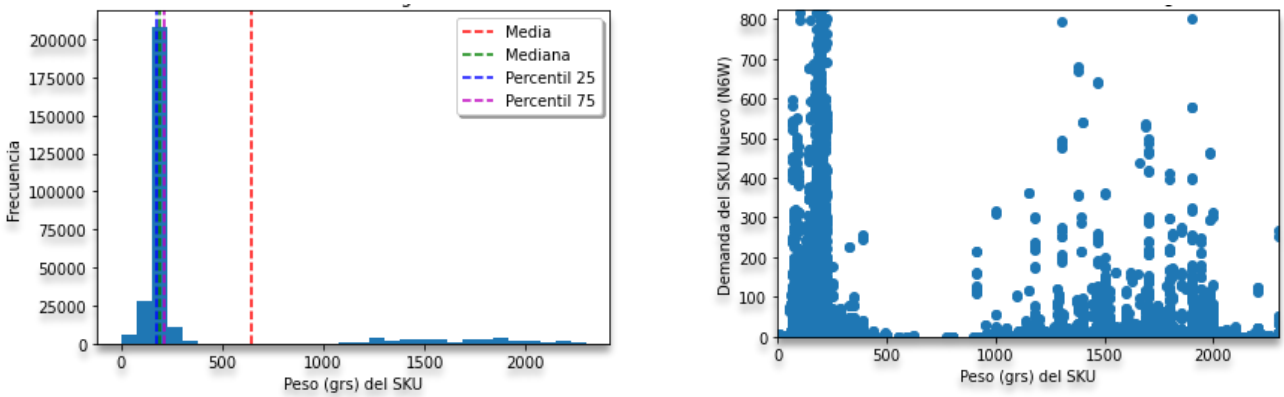


Figura 9.15 Izquierda: Distribución de observaciones según volumetría del SKU. Derecha: Distribución de la demanda según volumetría del SKU

La Figura 9.15 a la derecha indica que los SKUs nuevos con un peso de hasta 250 gramos tienen una mayor cantidad de ventas en las primeras seis semanas que aquellos que superan este peso, excepto por algunos casos particulares.

Al igual que la variable de volumetría, el peso está directamente relacionado con la categoría del producto (celulares o *notebooks*). Por lo tanto, no aporta información adicional significativa para entrenar al modelo.

9.5 *Feature Importance* de los Modelos Entrenados

9.5.1 *Modelo 1: Random Forest Regressor*

Al analizar la Figura 9.16 que muestra las diez *features* más relevantes para el modelo, se destaca que la variable *'atribute_qty'* (cantidad de atributos) sobresale como la más influyente. Esto sugiere que, para el modelo, la cantidad de atributos en la publicación es un factor crítico en la estimación de la demanda de un producto nuevo sin historial de ventas en las primeras seis semanas. Además, se observa que las variables relacionadas con el tamaño del producto (por ejemplo, *'depth_cm'*) y la reputación del vendedor, especialmente si pertenece a la categoría '*green platinum*' también influyen en las predicciones del modelo.

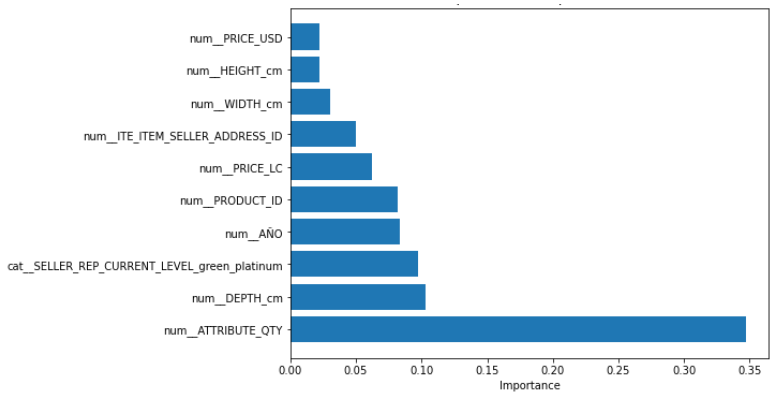


Figura 9.16 Top diez features más relevantes del modelo entrenado

9.5.2 *Modelo 2: XGBoost Regressor*

Al analizar la Figura 9.17 que detalla las características más relevantes para el modelo, se identifican diferencias con respecto al modelo *random forest regressor*. Si bien ambos modelos destacan una única característica como la más influyente, en el caso del *XGBoost regressor*, esta característica se relaciona con la reputación actual del vendedor, particularmente si pertenece a la categoría '*green*

platinum'. Esto contrasta con el modelo *random forest regressor*, para el cual la variable *'cant_atributos'* ocupa el primer lugar de relevancia.

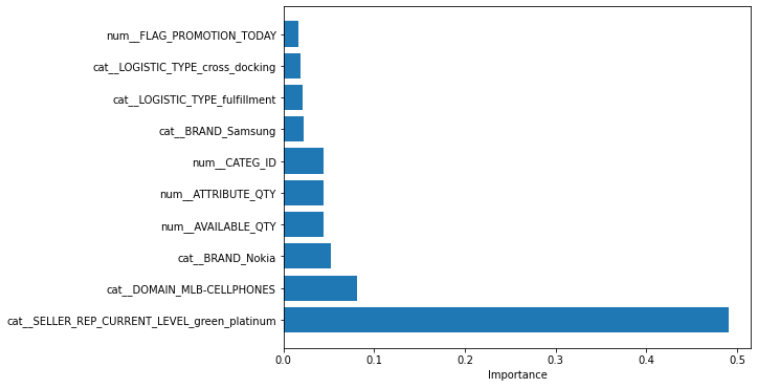


Figura 9.17 Top diez features más relevantes del modelo entrenado

9.5.3 *Modelo 3: Random Forest Classifier con Cuatro Clases*

En las Figuras 9.18 a 9.21, se exploran las *features* más importantes por clase. Todas las clases comparten las mismas top tres features: *'seller_reputation_current_level'* (nivel '*green_platinum'*) seguida de *'ventas historicas vendedor'* y '*available_qty'*. Esto sugiere que el modelo enfrenta dificultades para diferenciar entre clases por la influencia dominante de estas características comunes.

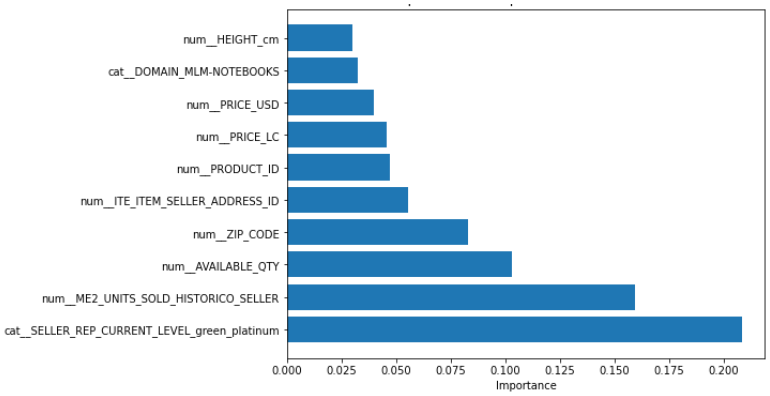


Figura 9.18 Top diez features más relevantes del modelo entrenado para la clase "0-10"

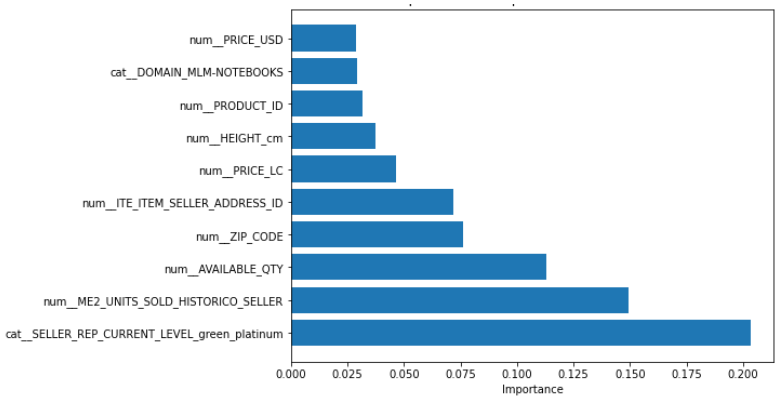


Figura 9.19 Top diez features más relevantes del modelo entrenado para la clase "11-20"

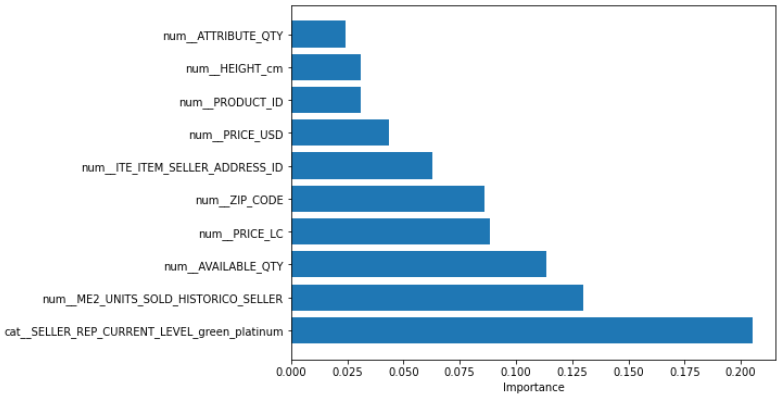


Figura 9.20 Top diez features más relevantes del modelo entrenado para la clase "21-50"

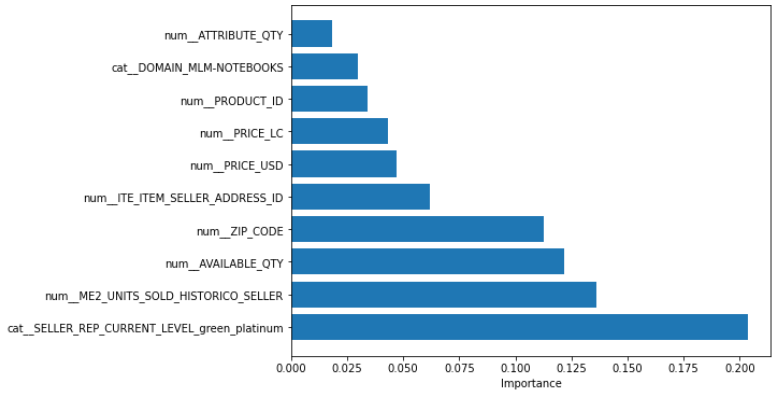


Figura 9.21 Top diez features más relevantes del modelo entrenado para la clase ">51"

9.5.4 *Modelo 4: XGBoost Classifier con Cuatro Clases*

Al analizar la importancia de las características más relevantes para el modelo, se observan diferencias notables en comparación con el modelo 3. Las Figuras 9.22 a 9.25 muestran que la característica *'Seller reputation current level'*, especialmente la categoría *'green platinum'*, sigue siendo la más influyente en todas las clases excepto para "11-20". Para esta clase, la característica más importante es *'Brand'*, particularmente *'Positivo'*. Además, se nota que la importancia relativa de *'Seller reputation*

current level' varía entre las clases. Por ejemplo, en la clase "0-10", aporta el 0.35 de la importancia, mientras que en las clases "21-50" y ">51", su contribución disminuye a aproximadamente 0.16 y 0.175.

En las clases "0-10" y "11-20", se resalta la importancia de características relacionadas con el dominio del producto y el historial de ventas del vendedor. En la clase "21-50", las características relacionadas con la logística, específicamente *fulfillment* y la marca *Xiaomi*, son relevantes. Por último, en la clase ">51", las características destacadas incluyen la logística, particularmente *'XD Drop off'*, y el historial de ventas del vendedor.

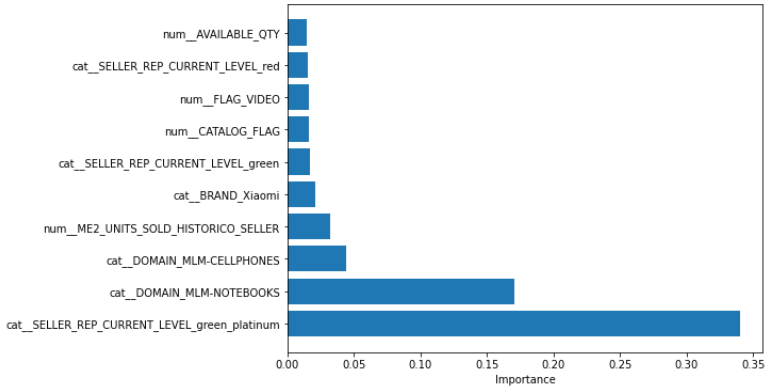


Figura 9.22 Top diez features más relevantes del modelo entrenado para la clase "0-10"

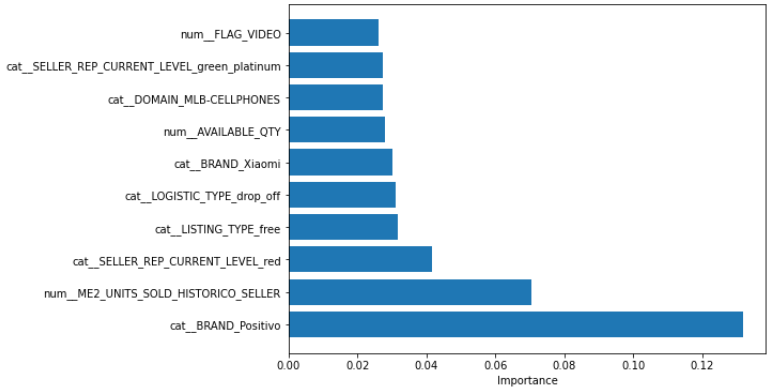


Figura 9.23 Top diez features más relevantes del modelo entrenado para la clase "11-20"

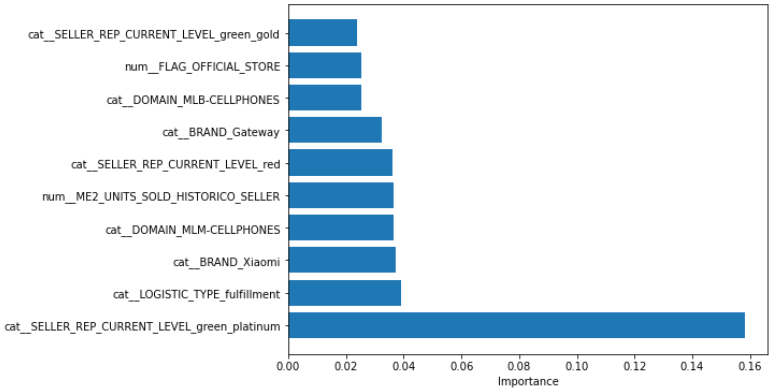


Figura 9.24 Top diez features más relevantes del modelo entrenado para la clase "21-50"

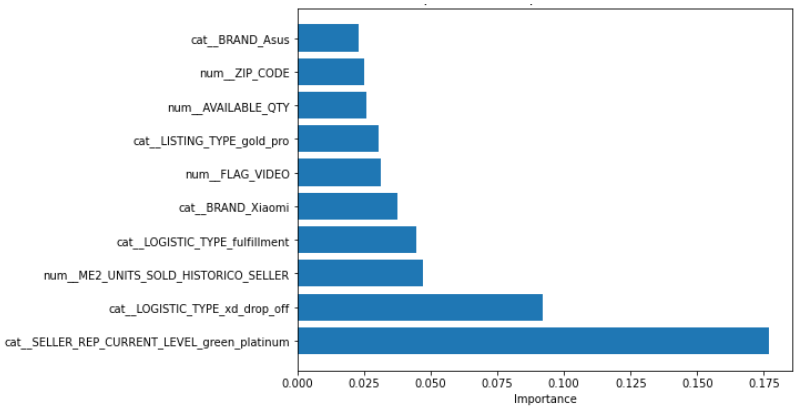


Figura 9.25 Top diez features más relevantes del modelo entrenado para la clase ">51"

9.5.5 *Modelo 5: Random Forest Classifier con Tres Clases*

Al analizar la importancia de las características para el modelo *random forest classifier* con 3 clases, se observan diferencias en comparación con los modelos anteriores.

Como se observa en la Figura 9.26, para la clase "0-10", la característica más relevante es *'Available_qty'*, seguida por el histórico de ventas del vendedor y el código postal. Todas las diez principales características tienen una relevancia bastante pareja entre ellas, aunque en general, su contribución es baja, llegando hasta un máximo de 0.08.

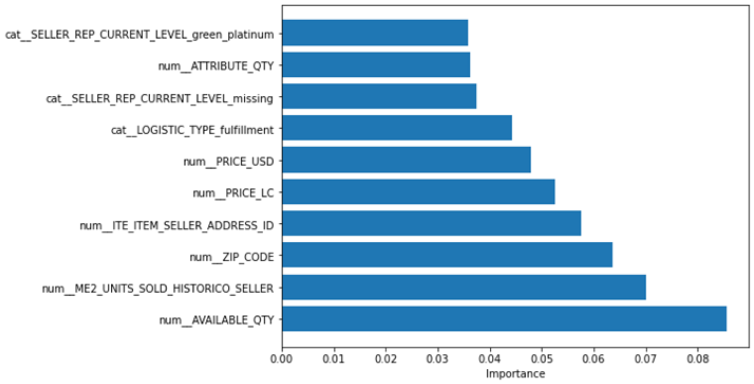


Figura 9.26 Top diez features más relevantes del modelo entrenado para la clase "0-10"

Para la clase "11-20", como muestra la Figura 9.27, la característica más relevante es *'Seller reputation level'*, particularmente la categoría *'green platinum'*, seguida por *'Available_qty'*. Nuevamente, estas características tienen baja contribución individual, con un máximo de 0.08.

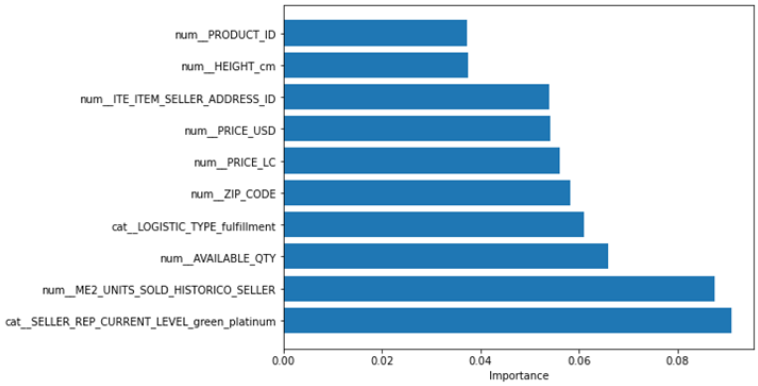


Figura 9.27 Top diez features más relevantes del modelo entrenado para la clase "11-20"

La Figura 9.28 explora la clase ">21", cuya característica más relevante es el histórico de ventas del vendedor, seguida por su ID de dirección y *'Available_qty'*. En este caso, la característica más importante tiene una contribución sustancial en relación con las demás, aportando un 0.16 de relevancia.

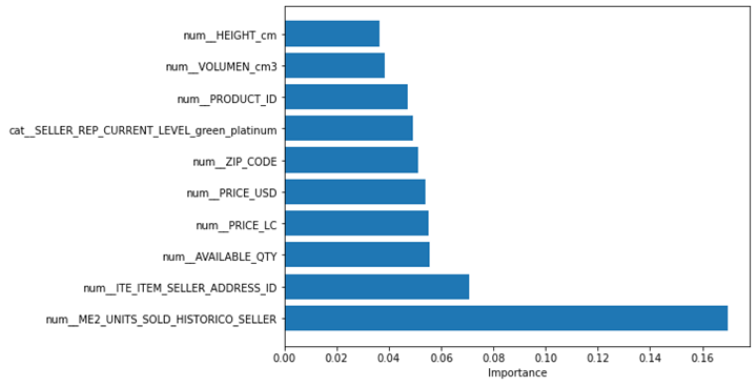


Figura 9.28 Top diez features más relevantes del modelo entrenado para la clase ">21"

9.5.6 *Modelo 6: XGBoost Classifier con Tres Clases*

El análisis de la importancia de las características para el modelo 6, un *XGBoost Classifier* con tres clases, revela diferencias notables en comparación con los modelos anteriores. En la clase "0-10", como se observa en la Figura 9.29, las características más relevantes son el dominio del producto, especialmente en México y en la categoría de *notebooks*, seguido por el nivel de reputación del vendedor, con énfasis en la categoría *'green_platinum'*. En tercer lugar, se destaca la marca del producto, particularmente '*Vivo*'. Las dos primeras características tienen una importancia significativa, con un peso de hasta 0.20, mientras que la marca tiene un peso de aproximadamente 0.075.

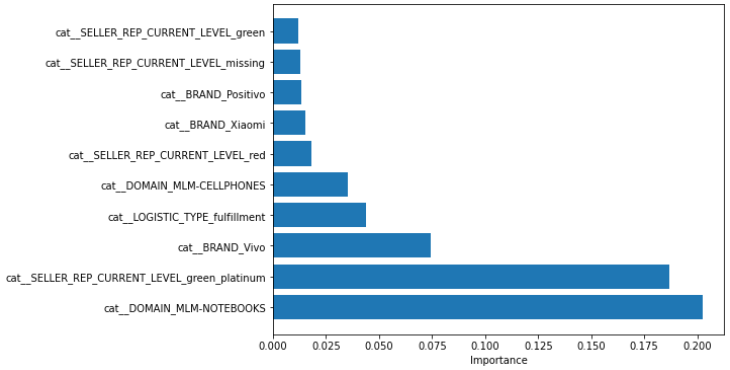


Figura 9.29 Top diez features más relevantes del modelo entrenado para la clase "0-10"

En la clase "11-20" explorada en la Figura 9.30, las diez características principales tienen relevancia más equilibrada y menor contribución individual, con un máximo de 0.04. La característica más importante es el tipo de publicación *'free'*, seguida por la marca *'Vivo'* y el dominio de México en la categoría *'cellphones'*.

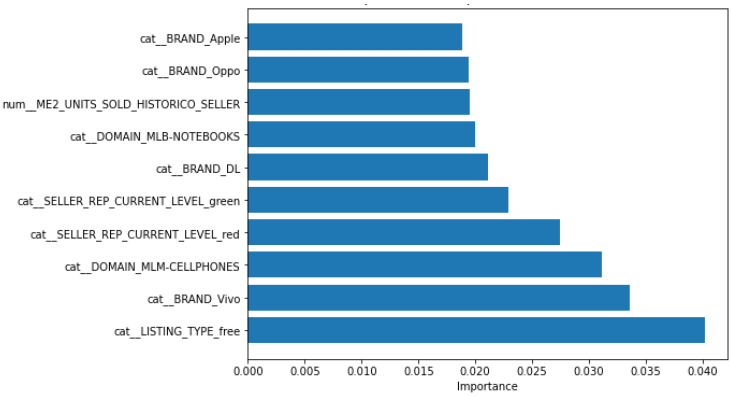


Figura 9.30 Top diez features más relevantes del modelo entrenado para la clase "11-20"

Para la clase ">21" analizada en la Figura 9.31, se observa un desequilibrio en la importancia de las características. La más relevante, con un peso de hasta 0.25, es la reputación del vendedor, específicamente si es *'green_platinum'"*. Le sigue la marca *'Gateway'* y el dominio de México en la categoría *'cellphones'*.

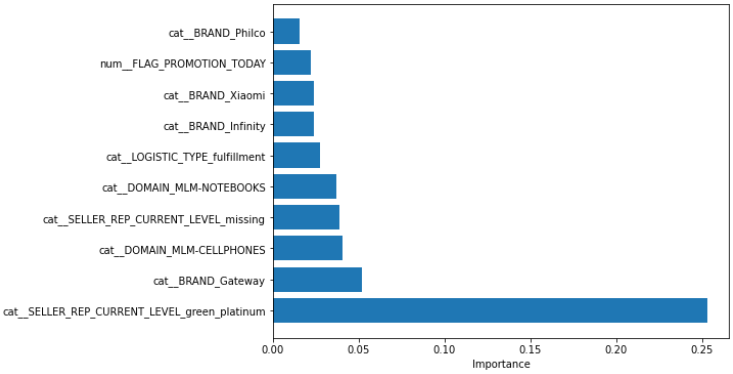


Figura 9.31 Top diez features más relevantes del modelo entrenado para la clase ">21"

[1] Peekpublishers (2023). *Role of Data Analytics in Logistics.* Logistics Peek.

[2] Marko. (2021). *E-commerce in Post-Pandemic World*. Younify.

[3] Shaw, N., Eschenbrenner, B., y Baier, D. (2022). *Online shopping continuance after COVID-19: A comparison of Canada, Germany, and the United States.* NCBI.

[4] Fiedler, L., Hazan, E., Ruwadi, B., y Ungerman, K. (2020). *Retail reimagined: The new era for customer experience*. [Archivo PDF].

[5] *We are Mercado Libre*. (s.f.). [Archivo PDF].

[6] Fatemi, Z., Huynh, M., Zheleva, E., Syed, Z., & Di, X. (2023). *Mitigating Cold-start Forecasting using Cold Causal Demand Forecasting Model.* [Archivo PDF].

[7] Coppola, D. (2022). *E-commerce as Share of Total Retail Sales Worldwide 2015-2021 with Forecasts to 2026.* Statista.

[8] (2021). *How Covid-19 triggered a Latin American e-commerce boom.* OxfordBusinessGroup.

[9] Stanley, L. (2022). *The Complete History of Ecommerce.* Nexcess.

[10] (2023). *The Evolution of Digital Payments and E-Commerce. FinanceMagnates*.

[11] Tian, Y. y Stewart, C. (2007). *History of e-commerce. In Electronic Commerce: concepts, methodologies, tools, and applications.* IGI Global.

[12] (2023). *Measuring the value of e-commerce.* UNCTAD. [Archivo PDF].

[13] Perrin, A., y Duggan, M. (2015). *American's Internet Access: 2000-2015.* PewResearch.

[14] Gitlan, D. (2024). *SSL History: From Inception to Evolutionary Triumph*. SSLDragon.

[15] *FedEx Annual Report* (2005). [Archivo PDF].

[16] *Global E-Commerce Market Size & Share Analysis - Growth Trends & Forecasts (2023 - 2028).* (s.f.). MordorIntelligence.

[17] (2021). *Machine Learning: AI's Role in the Future of eCommerce.* VARStreet.

[18] Oxford Learner's Dictionaries. (s.f.). Marketplace.

[19] Rodriguez, J. (2023). *Qué es un marketplace: ejemplos, tipos y cómo funcionan.* Hubspot.

[20] Joshua, J. M. (2022). *El comercio en la antigua Mesopotamia.* WorldHistory.

[21] *Mercaderes y Ferias en la Edad Media: Rutas Comerciales.* (s.f.). HistoriaYBiografias.

[22] *Historia del Dinero*. (s.f.). En Wikipedia.

[23] *El impacto del telégrafo en el comercio y la economía global.* (s.f.). Historioteca.

[24] (2023). *La historia de Amazon, el e-commerce gigante que revolucionó la industria*. TheLogisticsWorld.

[25] Naveira, A. (2020*). Historia de eBay: nacimiento y evolución de uno de los mayores marketplaces del mundo*. Marketing4Ecommerce.

[26] Pop, A. (2022). *Una guía sobre los diversos tipos de marketplaces.* Vtex.

[27] *What Is An Online Marketplace?* (s.f.). Stockarea.

[28] Ecker, T., Hans, M., Neuhaus, F., y Spielvogel, J. (2020). *Same-day delivery: Ready for takeoff.* McKinsey.

[29] *Beneficios de la membresía Prime.* (s.f.). Amazon.

[30] *Estadísticas de Amazon Prime 2023: ¿Cuántas personas usan Amazon Prime?* (s.f.). Pctg.

[31] Found, N., Lim, R., Holmes, J., y Duoanla, J. (2022). *Ecommerce Delivery Benchmark Report 2022. Welcome to the age of ecommerce*. [Archivo PDF].

[32] (2020). *What is Dropshipping? Benefits, Challenges, and Getting Started.* Adobe.

[33] Lopienski, K. (2023). *Cross-docking explained: Here's what you need to know*. Shipbob.

[34] Mehta, D. (2023). *Fulfillment by Amazon: How improving delivery fueled independent seller growth and success.* AboutAmazon.

[35] Seyedan, M., y Mafakheri, F. (2020). *Predictive big data analytics for supply chain demand forecasting: methods, applications, and research opportunities.* JournalOfBigData.

[36] *Demand Forecasting: 10 key complexity drivers.* (s.f.). Quantics.

[37] Thonemann, U., y Bradley, J. (2002). *The Effect of Product Variety on Supply-Chain Performance*. European Journal of Operational Research.

[38] Han, J., Kamber, M., y Pei, J. (2011). *Data mining: Concepts and techniques (3rd ed.)*. Morgan Kaufmann

[39] Han, J., Kamber, M., y Pei, J. (2011). *Data mining: Concepts and techniques (3rd ed.)*. Morgan Kaufmann

[40] Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.

[41] Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.

[42] Goodfellow, I., Bengio, Y., y Courville, A. (2016). *Deep Learning*. MIT Press.

[43] Goodfellow, I., Bengio, Y., y Courville, A. (2016). *Deep Learning.* MIT Press.

[44] Hastie, T., Tibshirani, R., & Friedman, J. (2009*). The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer.

[45] Hastie, T., Tibshirani, R., & Friedman, J. (2009*). The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer.

[46] Hastie, T., Tibshirani, R., y Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* Springer.

[47] Dieckmann, J. (2023). *Ensemble Learning: Bagging and Boosting.* Towards Data Science.

[48] Hastie, T., Tibshirani, R., y Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Science & Business Media.

[49] Hastie, T., Tibshirani, R., y Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* Springer Science & Business Media.

[50] Dieckmann, J. (2023). *Ensemble Learning: Bagging and Boosting.* Towards Data Science.

[51] Dieckmann, J. (2023). *Ensemble Learning: Bagging and Boosting.* Towards Data Science.

[52] Hyndman, R. J., y Koehler, A. B. (2006). *Another look at measures of forecast accuracy.* ScienceDirect.

[53] Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective.* MIT Press.

[54] *OneVsRestClassifier.* (s.f.). ScikitLearn.

[55] Bergstra, J. y Bengio, Y. (2012). *Random search for hyper-parameter optimization.* Journal of Machine Learning Research.

[56] Hastie, T., Tibshirani, R., & Friedman, J. (2009*). The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer.

[57] Raschka, S., y Mirjalili, V. (2019). *Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow 2.* Packt Publishing.

[58] Dietterich, T. G. (1995). *Overfitting and under-computing in machine learning.* ACM Computing Surveys.