

**Tipo de documento:** Tesis de maestría

Escuela de Negocios. Master in Management + Analytics

# Estimación de precios de mercado para celulares usados mediante técnicas de aprendizaje automático

**Autoría:** Rodríguez, Federico Martín

**Año:** 2024

## ¿Cómo citar este trabajo?

Rodríguez, F. (2024). "Estimación de precios de mercado para celulares usados mediante técnicas de aprendizaje automático". [Tesis de maestría. Universidad Torcuato Di Tella]. Repositorio Digital Universidad Torcuato Di Tella. <https://repositorio.utdt.edu/handle/20.500.13098/12786>

El presente documento se encuentra alojado en el Repositorio Digital de la **Universidad Torcuato Di Tella** bajo una licencia Creative Commons Atribución-Compartir igual 4.0 Internacional Deed

**Dirección:** <https://repositorio.utdt.edu>

# Estimación de precios de mercado para celulares usados mediante técnicas de aprendizaje automático

Master in Management + Analytics

Universidad Torcuato Di Tella

Alumno: Federico Martín Rodríguez

Tutor: Emmanuel Iarussi

# Resumen

En este trabajo se aborda el desafío de automatizar la toma de decisiones de una empresa poniendo foco en la satisfacción del cliente a través de ofertas instantáneas a precios de mercado que incentiven al usuario a volver siempre que desee actualizar su celular mediante la venta de su dispositivo usado. Para esto se propone un algoritmo que utiliza los datos recolectados de una plataforma de ecommerce con el objetivo de determinar el valor de mercado que tienen los dispositivos con el fin de asignar de manera automática la oferta a realizar por el celular usado del usuario.

En el algoritmo se incluyen las características asociadas no solamente con el dispositivo sino también con el vendedor y el tipo de publicación (reputación, garantías, etc...), diferenciando a los vendedores individuales de las empresas para obtener los resultados a considerar a la hora de implementar esta solución. Entre las diferencias se determina (entre otros) que un agente individual no tiene reputación, no otorga garantías y tiene un tiempo de espera máximo de venta igual a 1 día. Para la empresa se asume que tiene una reputación consolidada, otorga garantías y tiene un tiempo de espera de venta de hasta 30 días.

Los resultados obtenidos permiten estimar un promedio de ganancia porcentual bruta del 12,52% que aumenta hasta 43,09% si la empresa puede reacondicionar el celular antes de venderlo. Se resalta la importancia de evaluar con expertise técnico las predicciones individuales de los distintos modelos para determinar si las ofertas generadas son congruentes con el mercado o aparecen distorsionadas por una baja o mala representación en la información.

En este trabajo se consideran tres condiciones de estado de los dispositivos; "Nuevo", "Reacondicionado" y "Usado". Para futuros desarrollos, se espera poder añadir información más específica de la condición que tienen los dispositivos por medio de imágenes y descripciones que no pudieron ser incluidas en esta tesis.

# Abstract

In this work we address the challenge of automating decision-making of a business with a focus on customer experience through instant market-price offers that encourage users to return whenever they want to upgrade their cell phones by selling their used device. To achieve this, an algorithm that uses data collected from an e-commerce platform to determine the market value of devices is proposed in order to automatically assign an offer for the user's used cell phone.

The algorithm takes into account characteristics associated with the device, the seller, and the type of listing (reputation, guarantees, etc.), differentiating individuals from companies. For individuals, we assume they have no reputation, do not offer guarantees, and have a maximum sale waiting time of 1 day. For companies, we assume they have a consolidated reputation, offer guarantees, and have a sale waiting time of up to 30 days.

Our results show an estimate of an average gross profit margin of 12.52%, which increases up to 43.09% if the company can refurbish the cell phone before selling it. The importance of evaluating the individual predictions of different models with technical expertise is highlighted to determine if the generated offers are consistent with the market or distorted due to low or poor representation in the data.

This thesis considers three conditions for the devices: "New," "Refurbished," and "Used." For future developments, it is expected to add more specific information about the condition of the devices through images and descriptions that were not included in the main document.

# Índice

<b>Resumen.....</b>	<b>1</b>
<b>Abstract.....</b>	<b>2</b>
<b>Índice.....</b>	<b>3</b>
<b>1. Introducción.....</b>	<b>6</b>
1.1. Productos altamente tecnológicos.....	6
1.2. Reacondicionamiento.....	7
1.3. Depreciación.....	7
1.4. Mercado Argentino: oportunidad.....	7
1.5. Definición del problema.....	8
1.6. Solución propuesta.....	9
<b>2. Estado del arte.....</b>	<b>10</b>
<b>3. Metodología.....</b>	<b>13</b>
3.1. Recolección de fuente principal de datos vía API.....	13
3.2. Recolección de fuente secundaria de datos vía Beautiful Soup.....	14
3.3. Extracción de características de los títulos de las publicaciones vía modelos de lenguaje... 14	
3.4. Modelos de regresión.....	15
3.4.1. Regresión lineal.....	15
3.4.2. Árbol de decisión.....	16
3.4.3. Random forest.....	17
3.4.4. XGBoost (Extreme Gradient Boosting).....	18
3.4.4.1. Regularización L1 (Lasso).....	18
3.4.4.2. Regularización L2 (Ridge).....	18

3.5. Métricas de evaluación.....	19
3.5.1. Raíz del Error Cuadrático Medio (RMSE).....	19
3.5.2. Coeficiente de Determinación (R2).....	19
3.5.3. Error Porcentual Absoluto Medio (MAPE).....	20
3.6. Grado de ajuste del modelo.....	20
3.6.1. Subajuste.....	20
3.6.2. Sobreajuste.....	20
<b>4. Datos.....</b>	<b>22</b>
4.1. Recolección.....	22
4.2. Diseño.....	23
4.3. Análisis exploratorio de los datos recolectados.....	26
4.3.1. Precio de las publicaciones en el tiempo.....	30
4.3.2. Atributos de los celulares.....	33
4.3.3. Tags asociados a las publicaciones.....	34
4.3.4. Títulos de las publicaciones.....	35
4.3.5. Atributos estandarizados de GSMArena.....	38
4.4. Preprocesamiento.....	42
4.4.1. Integración de las fuentes de datos.....	42
4.4.2. Ajuste por inflación de precios de publicaciones.....	44
4.4.3. Identificación y eliminación de registros problemáticos, separación del dataset en conjuntos de entrenamiento y evaluación.....	45
4.4.4. Transformación de la variable objetivo.....	47
<b>5. Resultados y análisis de las predicciones.....</b>	<b>50</b>
5.1. Evaluación de modelos base.....	50

5.1.1. Test de Diebold-Mariano para la evaluación de la dominancia de un modelo.....	58
5.2. Evaluación de modelos ajustados.....	59
5.2.1. Búsqueda aleatoria de hiperparámetros con validación cruzada.....	59
5.2.2. Hiperparámetros de los modelos.....	61
5.2.3. Resultados de los modelos ajustados.....	62
5.2.4. Importancia de los atributos del modelo.....	67
5.2.5. Comparación contra modelo individual de dispositivo.....	68
<b>6. Consideraciones generales.....</b>	<b>76</b>
6.1. Supuestos subyacentes en los datos.....	76
6.2. Extensiones.....	76
6.3. Implementación propuesta.....	77
<b>7. Conclusiones.....</b>	<b>79</b>
<b>Referencias.....</b>	<b>81</b>
<b>Anexo.....</b>	<b>83</b>

# 1. Introducción

Durante los últimos años la extensión de la vida útil de los dispositivos de telefonía móvil a través de la reventa aparece creciendo a razón de 4 a 5 veces más que el mercado de celulares en general, con ventas estimadas de 120 millones de unidades por un valor total de 17 mil millones de dólares en el mundo en 2016 (Makov et al. 2019). Una prueba de ello es el caso de Amazon Warehouse<sup>1</sup> donde se encuentran, entre las opciones disponibles, celulares reacondicionados y usados que no pueden ser catalogados como nuevos dados los estándares de la plataforma.

Este mercado, además de permitir a los consumidores ahorrar a la hora de obtener un celular funcional, promueve el reciclaje a través de la economía circular, haciéndolo una opción más sustentable ya que reduce el desperdicio electrónico, categoría de desechos sólidos peligrosos con más rápido crecimiento en el mundo (Awasthi et al. 2019).

Por otro lado, los celulares usados presentan poca o nula garantía, que se puede traducir en un costo para el comprador en el caso de presentarse alguna falla, riesgo que se ve incrementado a medida que aumenta la incertidumbre que se tiene con el vendedor. En las secciones siguientes, se enumeran las características de estos productos, las consideraciones que tienen sus consumidores, la oportunidad que se presenta en Argentina, los desafíos que conllevan y se proponen algunas soluciones.

## 1.1. Productos altamente tecnológicos

En contraste con los bienes de consumo duraderos, los teléfonos móviles tienen un ciclo de vida corto debido a la rápida obsolescencia tecnológica y la constante innovación en la industria. Siendo la empresa Apple<sup>2</sup> un claro ejemplo de esto, con al menos un lanzamiento nuevo cada año.

Esto resulta en la coexistencia de múltiples generaciones que compiten para satisfacer las preferencias de los consumidores, donde se encuentran por un lado aquellos que dan prioridad a la innovación y la adquisición de los modelos más recientes, y por el otro, quienes valoran más la calidad y el precio. En este contexto, los productos refabricados aparecen para extraerle aún más valor a los usados, llegando a influir en el valor de los dispositivos nuevos. La tendencia es hacia el aumento cuando quien realiza el reacondicionamiento aparece como tercera parte

---

<sup>1</sup> [Amazon Warehouse, 2023](#)

<sup>2</sup> [Apple, 2023](#)

frente a la empresa desarrolladora del dispositivo, y a disminuirlo cuando es la empresa quien se hace cargo del reacondicionamiento (Zhou et al. 2019).

## 1.2. Reacondicionamiento

En lo que a celulares refabricados respecta, se entiende que los consumidores hacen una evaluación acerca del valor y los riesgos asociados al reacondicionamiento del producto donde consideran su calidad, costo y sustentabilidad (Wang et al. 2016). En este sentido, la reputación construida por el vendedor provee de señales en lo que al servicio y la calidad del producto conlleva, ayudando a mitigar la incertidumbre de los consumidores al adquirir los productos refabricados. Cabe resaltar que garantías más fuertes no se asocian significativamente a una mayor disposición de pago por parte de los clientes (Subramanian et al. 2012).

## 1.3. Depreciación

Entendiendo la depreciación como la pérdida de valor que tienen los celulares en el intervalo de tiempo comprendido entre el lanzamiento hasta la reventa, podemos encontrar diversos estudios que detallan los factores que influyen en su aceleración y desaceleración. En este sentido, factores intangibles como la marca demuestran ser los más relevantes a la hora de resguardar el valor del producto a lo largo del tiempo cuando se la compara con factores tangibles como su condición, la capacidad (medida en gigabytes) y la reparabilidad, medida en la puntuación asignada por la comunidad mundial de reparación iFixit<sup>3</sup> (Makov et al. 2019). Se resalta que la marca está asociada a factores tangibles como la obsolescencia programada o vida útil que las empresas le asignan a sus teléfonos. Otros factores como el paso del tiempo, el espesor y el peso, aparecen acelerando la pérdida de valor que obtienen los dispositivos en el mercado secundario (Zhou et al. 2020).

## 1.4. Mercado Argentino: oportunidad

En Argentina el mercado en línea de celulares usados pareciera estar todavía dominado por los agentes individuales que utilizan sitios de compraventa como Mercado Libre<sup>4</sup> y Facebook Marketplace<sup>5</sup>. Y en estos lugares, los agentes se exponen a posibles problemas como:

- Problemas de calidad: Los productos pueden no cumplir con las expectativas de calidad del comprador, quien puede incluso quedar expuesto a posibles actividades fraudulentas (como la compra de un dispositivo robado).

---

<sup>3</sup> [iFixit, 2023](#)

<sup>4</sup> [Mercado Libre, 2023](#)

<sup>5</sup> [Facebook Marketplace, 2023](#)

- Problemas de devolución o reembolso: El desacuerdo entre comprador y vendedor sobre la devolución de un artículo o reembolso de dinero puede llevar a complicaciones y disputas.
- Problemas de precio: Tanto el comprador como el vendedor pueden encontrar dificultades a la hora de negociar un precio acorde al celular usado.
- Problemas de pago: En general, las opciones de pago se ven limitadas al efectivo o transferencia bancaria, donde aparece la posibilidad de sufrir estafas.
- Problemas de logística: Los envíos pueden retrasarse o perderse durante el transporte, deviniendo en una insatisfacción para el comprador o pérdida de dinero para el vendedor.
- Problemas de verificación de identidad: En algunas plataformas, la falta de una adecuada verificación de identidad puede facilitar la actividad fraudulenta.
- Problemas de seguridad: Encontrarse en persona para completar la venta presenta un costo de exposición en materia de seguridad.

Mitigar estos riesgos facilitando la experiencia de aquellos que anhelan actualizar su celular aparece como una oportunidad para las empresas dispuestas a adoptar la compraventa de celulares. Donde la experiencia que tiene el usuario tiene el potencial de convertirlo en parte del ecosistema, incentivando la recurrencia cada vez que necesite cambiar su dispositivo e impulsando las ventas de usados gracias a la confianza conseguida.

## 1.5. Definición del problema

Dar crédito al cliente a cambio de su celular conlleva algunos desafíos para las empresas del sector:

- Evaluación de valor justo: Determinar el valor justo del usado es un desafío ya que depende de factores como la marca, el modelo, la antigüedad, el estado físico y funcional, y la demanda en el mercado de segunda mano.
- Depreciación: Los celulares tienden a depreciarse rápidamente en valor, lo que significa que la empresa puede perder dinero en dispositivos usados sin valor de reventa.
- Reacondicionamiento: Los teléfonos pueden requerir de un reacondicionamiento antes de poder ser puestos a la venta en el mercado.

- Costos de almacenamiento y gestión: Almacenar, evaluar y gestionar los celulares requiere de recursos y espacio.

Estas problemáticas llevan a que, en la práctica, la aceptación del usado como parte de pago se limite a los últimos modelos, con valores desventajosos para el cliente, que también debe incurrir en los costos y riesgos de logística (ya que se realizan de manera presencial).

## 1.6. Solución propuesta

El objetivo de este proyecto es desarrollar un modelo basado en técnicas de *machine learning* capaz de predecir el valor de mercado de venta de un celular, utilizando las especificaciones técnicas, las variables relacionadas a las publicaciones y el tiempo de venta. De esta forma, la empresa podrá estimar la potencial rentabilidad asociada a su capacidad de espera en el tiempo, para ofrecer de manera instantánea una cotización correspondiente a la marca, modelo, capacidad y condición del celular del cliente.

Para esto se recopiló información diaria de las publicaciones de celulares en venta de Capital Federal dentro de Mercado Libre, plataforma digital de comercio electrónico líder en Argentina que conecta a vendedores con compradores. Las ventas se midieron con los cambios en las cantidades asociadas a los posteos, asumiendo que cualquier publicación con sólo una unidad disponible que desaparezca del sitio representa una venta.

## 2. Estado del arte

En la literatura sobre la predicción de precios de teléfonos móviles basada en sus características, es común observar que los estudios se enfocan en problemas de clasificación donde la principal tarea no es predecir el precio, sino la categoría a la que pertenece un dispositivo en función de su precio. En estos trabajos, encontramos que elegir una buena combinación de algoritmos, preprocesamiento y postprocesamiento de los datos tiene una gran importancia para conseguir buenos resultados (Çetin et al. 2022). Entre las investigaciones realizadas con esta metodología, podemos mencionar autores que utilizaron como fuente de datos GSMarena<sup>6</sup>, que es también utilizada en este trabajo para extraer las especificaciones de los celulares (detallado en la Sección 4.3.5), determinando la importancia de una buena selección de atributos (Asim et al. 2018) y mostrando mejoras en los resultados mediante la adición de atributos gráficos de los dispositivos (Zehtab-Salmasi et al. 2023).

Aunque en el ámbito de la regresión de atributos la literatura es escasa, podemos encontrar que se han llevado a cabo algunas investigaciones con un objetivo similar al de esta tesis: desarrollar modelos predictivos de precios de teléfonos celulares. A continuación, se detallan algunos de los trabajos más recientes en orden cronológico.

En Chandrashekhara et al. 2019, los autores entienden que el precio de los productos es el aspecto más importante a determinar por las empresas, asumiendo que es lo primero que mira el comprador. Los autores además suponen que a pesar de tener menos especificaciones y una calidad promedio, se puede conseguir que el celular sea exitoso en ventas. Frente a los métodos tradicionales utilizados para fijar el precio, proponen la utilización de modelos de machine learning entrenados con 262 registros recolectados de páginas de e-commerce que van desde marzo de 2010 hasta enero de 2018 (donde 210 se utilizan como entrenamiento y 52 como evaluación). Los registros contienen el precio de los celulares y sus especificaciones: marca, sistema operativo, RAM, memoria, peso, dimensiones, pantalla, tecnologías, cámaras primaria y secundaria, procesador, color y batería. Compararon diferentes modelos tradicionales como las Máquinas de Vector Soporte (SVM), Redes Neuronales y Regresión Lineal. En los experimentos realizados, las SVM reportaron la mejor performance, mientras que la regresión lineal estuvo entre los peores. Es importante destacar que este trabajo no detalla ningún ajuste inflacionario de los precios que, a nuestro parecer, es relevante dada la temporalidad que se tiene en los registros.

---

<sup>6</sup> [GSMarena.2023](#)

En Liu et al. 2020, los autores encuentran que la fijación automática de precios de los celulares reciclados está atrayendo cada vez más atención debido al crecimiento que tuvo el mercado en China. Luego de comparar distintos modelos de redes neuronales con distintas configuraciones, y destacando las relaciones de mapeo no lineales y conceptos de lógica difusa presentes en el problema, proponen una Red Neuronal Difusa con momento en la actualización de parámetros para una mejor fijación de precios. Cuentan con 1.200 registros de transacciones reales provenientes de compañías de reciclaje de teléfonos, pero en su preprocesamiento encuentran 128 valores atípicos, por lo que terminan utilizando los primeros 1.000 datos como conjunto de entrenamiento y los últimos 72 como conjunto de evaluación. Determinando que los atributos clave que afectan el precio son la marca, modelo, canal de compra, color, versión, almacenamiento, estado de garantía, apariencia trasera y de pantalla, rendimiento de pantalla y desmontaje.

En Maesya et al. 2023 dan respuesta a la falta de estudios sobre modelos de fijación de precio de los celulares (ya que la mayoría de las investigaciones existentes predicen las clases de precios de los teléfonos). En este trabajo proponen la comparación de dos algoritmos populares de aprendizaje supervisado para predecir su valor real. Por un lado utilizan el algoritmo de Bosque Aleatorio, que promedia los resultados de múltiples árboles de decisión con una selección aleatoria de características, utilizado con éxito para la predicción de, entre otras cosas, precios de casas y acciones. Por otro lado, utilizan la Regresión Lineal, que predice un valor objetivo creando una función lineal de los valores de entrada, utilizado también para predecir precios de, entre otras cosas, casas, acciones, oro, etc. La muestra de datos utilizada consta de 145 precios de celulares junto con sus características, las cuales son la cantidad de ventas, peso, resolución de pantalla, cantidad de píxeles, núcleos de CPU, frecuencia de CPU, tamaño de memoria interna, tamaño de RAM, resolución de la cámara trasera y frontal, capacidad de la batería y grosor del teléfono. Utilizando dos escenarios de distribución de datos (80:20 y 70:30) determinan que la resolución y el peso no están significativamente relacionados con el precio y que ambos modelos presentan capacidades similares con resultados de MAPE por debajo del 10% y  $R^2$  arriba del 94%.

En Liang et al. 2024 se comparan, al igual que en esta tesis, cuatro modelos de regresión asumiendo que los precios de los teléfonos móviles pueden predecirse con precisión utilizando algoritmos de aprendizaje automático. En el estudio utilizan Regresión Lineal, Bosque Aleatorio, XGBoost y Máquinas de Vector Soporte en base sus resultados en el  $R^2$ . Los datos utilizados consisten de 161 registros con el precio de cada celular, sus ventas, peso, resolución, densidad de píxeles, tipo de CPU, frecuencia del CPU, memoria interna, RAM, cantidad de cámaras traseras

y frontales, capacidad de la batería y grosor. Los datos se distribuyen en 112 filas para entrenamiento y 49 para evaluación, que representan el 70% y 30% de los datos respectivamente. Con un  $R^2$  de casi 95% se determina que el modelo de XGBoost es el que mejor ajusta a los datos, mientras que el de Máquinas de Vector Soporte, con un  $R^2$  de menos del 77%, el que menos logra ajustarse. También se concluye que la memoria interna es el factor más significativo a la hora de determinar el precio de los celulares.

A pesar del creciente interés en la automatización de la determinación del precio de los teléfonos celulares, pareciera que todavía nos encontramos aún en una fase incipiente de desarrollo de este problema. Los estudios reportados se enfocan en contextos en donde por lo menos una de las partes involucradas en la transacción es una empresa que posee una clara ventaja comparativa sobre el individuo que está del otro lado, sin considerar a los agentes presentes en el mercado secundario. Explorar el mercado de reventa nos aporta otra mirada al problema, que permite comprender de manera más exhaustiva el proceso de determinación del precio de los dispositivos donde no solo se consideran los atributos del teléfono celular sino también los correspondientes al ecosistema y los tiempos que conllevan una transacción.

### 3. Metodología

En esta sección se desarrollan las especificaciones de las técnicas utilizadas en este trabajo, desde la limpieza de los datos, hasta los modelos utilizados y su evaluación.

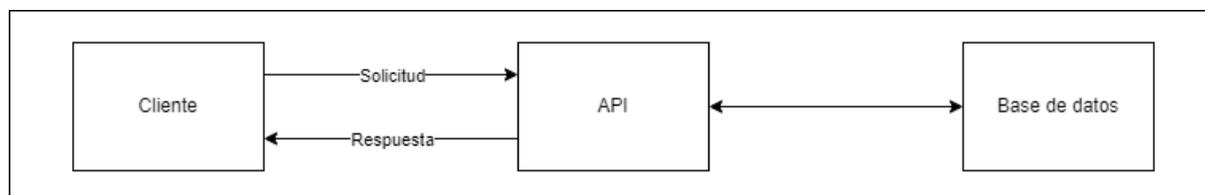
#### 3.1. Recolección de fuente principal de datos vía API

Los datos son la materia prima utilizada para extraer información, conocimiento y, como objetivo final, tomar decisiones basadas en evidencia acerca del precio a asignar a los dispositivos. Como principal fuente de este trabajo se utilizaron publicaciones de celulares nuevos, reacondicionados y usados, publicados en Mercado Libre tanto por empresas como por individuos.

Este acceso a los datos se hizo mediante la API de Búsqueda<sup>7</sup> y API de Producto<sup>8</sup> que la plataforma provee. Una API o interfaz de programación de aplicaciones es un conjunto de definiciones y protocolos que se utilizan para diseñar e integrar el software de las aplicaciones, determinando la manera y los tipos de solicitudes que los programadores pueden enviar y recibir (Bloch 2006).

Es importante tener en cuenta que las interfaces de programación de aplicaciones no son una tecnología específica, sino un concepto general que se aplica a muchas tecnologías diferentes. Estas pueden ser diseñadas para trabajar con una variedad de lenguajes de programación y plataformas, lo que las hace muy flexibles y adaptables a diferentes entornos de desarrollo. En términos simples, como se ilustra en la Figura 1, una API es un intermediario que nos permite comunicarnos con el servidor.

Figura 1. Ilustración del funcionamiento de una API.



<sup>7</sup> [API de Búsqueda de Mercado Libre, 2023](#)

<sup>8</sup> [API de Producto de Mercado Libre, 2023](#)

## 3.2. Recolección de fuente secundaria de datos vía BeautifulSoup

Además de la extracción de datos mediante la API, se requirió de una extracción adicional de información pertinente a los dispositivos móviles ya que, como desarrollaremos dentro de la Sección 4, sus especificaciones no están estandarizadas dentro de la plataforma, encontrando modelos con más de una asignación de atributo de único valor como por ejemplo la marca. Esta fue obtenida mediante la librería de python BeautifulSoup, que permite extraer información de contenido en formato HTML o XML.

Para usarla es necesario especificar un parser que es responsable de transformar un documento HTML o XML, en un árbol complejo de objetos Python. Esto permite, por ejemplo, que podamos interactuar con los elementos de una página web como si estuviésemos utilizando las herramientas del desarrollador de un navegador.

Para hacerlo, debemos seguir los siguientes pasos:

1. Identificar los elementos de la página de los que extraer la información.
2. Descargar el contenido de la página.
3. Crear la “sopa” de información.
4. Buscar los elementos y obtener la información deseada.

## 3.3. Extracción de características de los títulos de las publicaciones vía modelos de lenguaje

Los modelos de lenguaje son sistemas computacionales diseñados para entender, interpretar y generar lenguaje natural, estos forman parte del campo de estudio conocido como Procesamiento del Lenguaje Natural (NLP) y se utilizan para tareas como traducción automática, resumen de texto, generación de texto y más. Estos aprenden patrones y estructuras lingüísticas a partir de grandes cantidades de datos textuales en su fase de pre entrenamiento. Luego, pueden ser afinados para tareas específicas mediante la transferencia de aprendizaje, lo que significa que puede ser adaptado para tareas más específicas, como la extracción de datos con conjuntos de datos más pequeños y específicos.

En el contexto de la extracción de entidades nombradas (NER) relevante para nuestro caso, los modelos de lenguaje pueden ser utilizados para analizar grandes cantidades de texto, identificar patrones y extraer información relevante no sólo en función de las palabras anteriores, sino

también de las palabras posteriores en el título de la publicación. Por este motivo, se decidió utilizar el modelo de lenguaje “text-bison@001” de la familia de modelos PaLM 2<sup>9</sup> para extraer información adicional de los títulos de las publicaciones y así corregir valores faltantes o mal asignados que tengamos en los registros.

## 3.4. Modelos de regresión

Un modelo de regresión es un algoritmo que busca establecer una relación matemática entre variables independientes (también llamadas predictores) y una variable dependiente (la que queremos predecir), siendo su objetivo encontrar una función que permita estimar el valor de la variable dependiente en función de los valores de las variables independientes. En este trabajo exploramos las capacidades de los modelos de Regresión Lineal, Árboles de Decisión, Random Forest y XGBoost sobre los datos correspondientes a las publicaciones en venta de los celulares para la predicción de su precio de venta.

### 3.4.1. Regresión lineal

La regresión lineal es una técnica que se utiliza para modelar la relación entre una variable dependiente y una o más variables independientes donde se asume que su relación es lineal, lo que significa que los cambios en las variables independientes están asociados con cambios lineales en la variable dependiente.

$$\hat{y}(w, x) = \hat{w}_0 + \hat{w}_1 x_1 + \dots + \hat{w}_n x_n.$$

En su ecuación, denotamos  $\hat{y}$  como el valor predicho,  $\hat{w}_0$  como la constante (el valor que adopta la predicción cuando las variables independientes son iguales a cero) y a los vectores  $\hat{w} = (\hat{w}_1, \dots, \hat{w}_n)$  y  $x = (x_1, \dots, x_n)$  como los coeficientes estimados y los valores de las variables independientes.

Entre los supuestos del modelo encontramos: linealidad en la relación, independencia de los errores, homocedasticidad (varianza constante de los errores), normalidad de los errores, ausencia de multicolinealidad entre las variables independientes y exogeneidad (media condicional del error es cero).

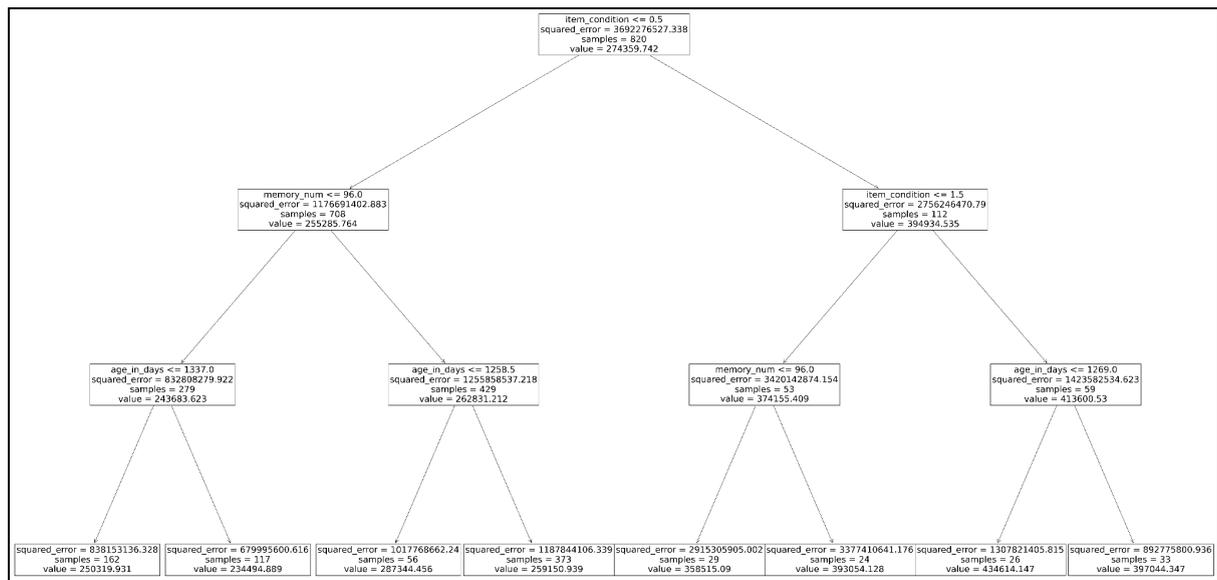
---

<sup>9</sup> [PaLM 2. 2023](#)

### 3.4.2. Árbol de decisión

En el contexto de problemas de regresión, el objetivo principal de los árboles de decisión es predecir un valor numérico basándose en un conjunto de características (atributos) que describen un objeto o una situación, estos modelos se llaman "árboles" debido a que su estructura visual es similar a la de un árbol invertido, como se observa en la Figura 2.

Figura 2. Ilustración de un árbol de decisión entrenado para predecir el precio de un iPhone 11 a partir de un subconjunto de atributos.



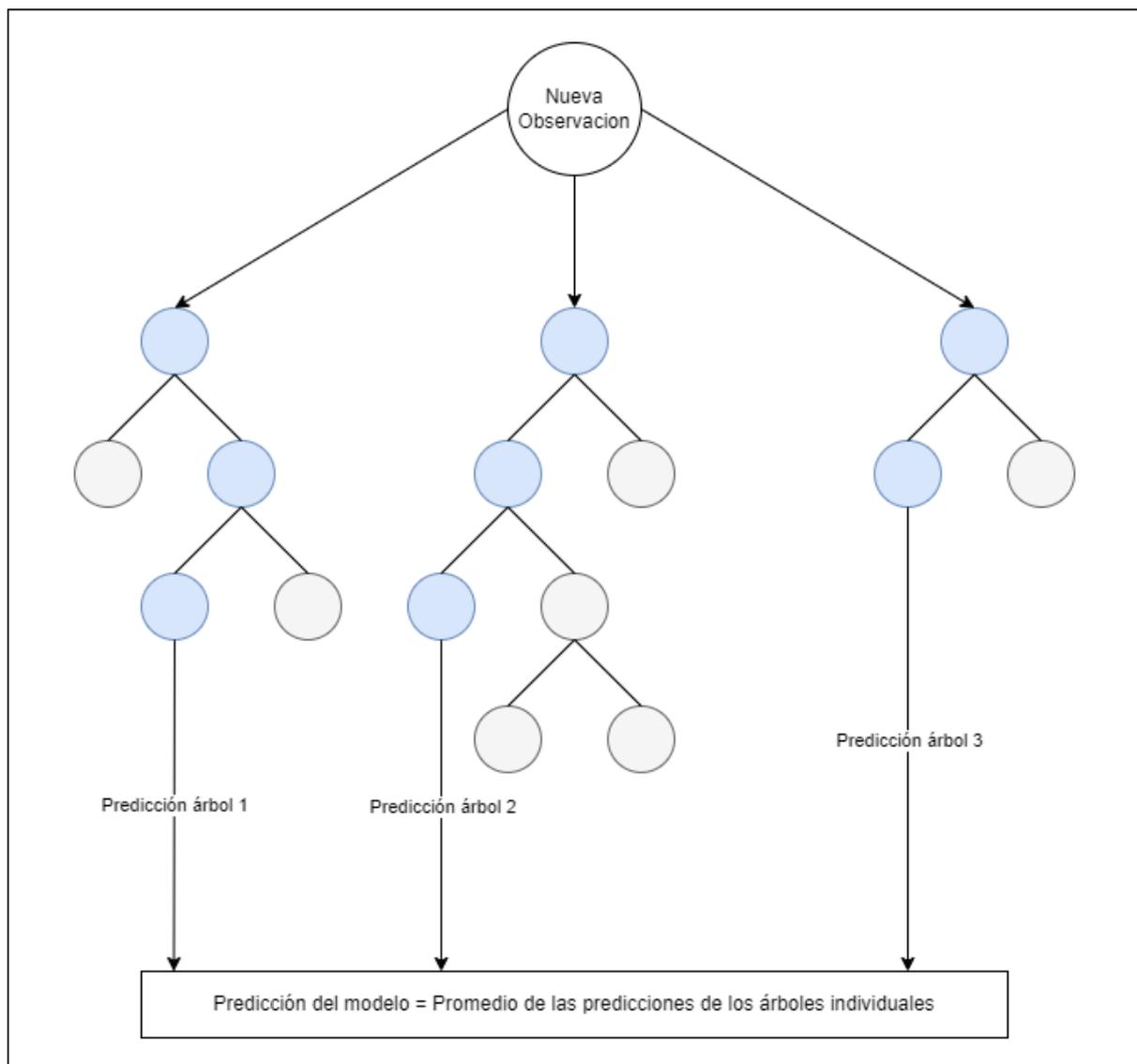
El proceso de creación de estos árboles comienza seleccionando un atributo del conjunto de datos que se utilizará para tomar la primera decisión, la elección del atributo se basa en su capacidad para dividir efectivamente los datos en grupos homogéneos en términos de la variable objetivo que se está prediciendo. Una vez que se ha seleccionado un atributo, se divide el conjunto de datos en subconjuntos más pequeños según los valores de ese atributo. Cada subconjunto se convierte en un nodo en el árbol. El mismo proceso de selección de atributos y división de nodos se repite en cada nodo hijo, creando así ramas adicionales en el árbol hasta que se cumple una condición de parada, como un número máximo de niveles en el árbol o la homogeneidad suficiente en los nodos.

Entre sus ventajas, los árboles de decisión resultan fáciles de entender y visualizar, tienen la capacidad de manejar tanto datos numéricos como categóricos y no requieren de la normalización de los datos. Entre las desventajas vemos que pueden ser propensos al sobreajuste (*overfitting*) si se permiten árboles muy profundos, también, la construcción del árbol puede ser sensible a pequeñas variaciones en los datos de entrenamiento, lo que los convierte en modelos menos precisos si los comparamos con técnicas más avanzadas.

### 3.4.3. Random forest

Desarrollado como una mejora de los árboles de decisión, su objetivo principal es mejorar la precisión y reducir el sobreajuste que presentan los árboles de decisión simples. Se basa en un conjunto de árboles de decisión donde cada uno toma decisiones basadas en reglas simples, dividiendo el conjunto de datos en función de características específicas para predecir una variable objetivo (Figura 3).

Figura 3. Ilustración del proceso predictivo de un random forest de regresión.



Cada árbol se construye utilizando una muestra aleatoria de los datos de entrenamiento y una selección aleatoria de las características, lo que introduce variabilidad en la construcción de los árboles, evitando que un solo atributo domine la predicción y promoviendo la diversidad entre los árboles. Al final, en el caso de regresión, se promedian las predicciones de todos los árboles para obtener la predicción final.

### 3.4.4. XGBoost (Extreme Gradient Boosting)

En lugar de entrenar árboles de decisión independientes, este modelo construye árboles de decisión secuencialmente de manera iterativa, donde cada uno se enfoca en corregir los errores del árbol anterior, lo que generalmente resulta en modelos más precisos. Además, XGBoost incorpora la regularización en la construcción de árboles, ayudando a prevenir el sobreajuste y aumentando la generalización del modelo.

La regularización en la construcción de modelos se utiliza para prevenir el sobreajuste de un modelo mediante la adición de términos adicionales a la función de costo (*loss function*) durante el proceso de entrenamiento. Estos términos adicionales penalizan los coeficientes de las características del modelo, evitando que tomen valores extremadamente grandes.

#### 3.4.4.1. Regularización L1 (Lasso)

En esta técnica, se agrega un término de penalización a la función de costo que se calcula tomando el valor absoluto de los coeficientes de las características y multiplicándose por un hiper parámetro de regularización lambda ( $\lambda$ ), como se detalla en la Sección 5.2.1, en este trabajo el valor de lambda será asignado mediante la búsqueda aleatoria con validación cruzada. Como el modelo intenta minimizar una nueva función de costo que reduce las pendientes de manera simultánea, algunos coeficientes cuyos pesos son más pequeños pueden llegar a volverse exactamente cero, haciendo que las características sean completamente ignoradas por el modelo ya que no son informativas.

Siendo  $w$  los coeficientes del modelo:

$$L1 \text{ Penalty} = \lambda \times \sum |w|.$$

#### 3.4.4.2. Regularización L2 (Ridge)

La penalización L2 se calcula tomando el cuadrado de los coeficientes de las características y multiplicándolos por  $\lambda$ , penalizando los coeficientes grandes de manera más suave que con la técnica de penalización L1 ya que tiende a forzar los coeficientes hacia valores pequeños en lugar de cero.

$$L2 \text{ Penalty} = \lambda \times \sum w^2.$$

## 3.5. Métricas de evaluación

Las métricas de evaluación son herramientas cuantitativas para medir y comparar el rendimiento de los distintos modelos, ayudando a tomar decisiones informadas sobre cuál modelo utilizar. Como muchas de estas métricas miden de manera distinta los resultados obtenidos por un modelo, utilizaremos tres de las más comunes dentro de los problemas de regresión para determinar cuál es el mejor para nuestro caso.

### 3.5.1. Raíz del Error Cuadrático Medio (RMSE)

Se calcula tomando la diferencia entre las predicciones del modelo y los valores reales, elevando esta diferencia al cuadrado, promediando por la cantidad de observaciones, y aplicando la raíz cuadrada. Por lo tanto, un valor más bajo de RMSE implica un mejor rendimiento del modelo.

Lo definimos como:

$$RMSE = \sqrt{\frac{\sum(y_i - \hat{y}_i)^2}{n}},$$

donde  $y_1, \dots, y_n$  son los valores reales,  $\hat{y}_1, \dots, \hat{y}_n$  las predicciones y  $n$  el número de observaciones.

### 3.5.2. Coeficiente de Determinación ( $R^2$ )

Mide la proporción de la varianza en la variable dependiente que es explicada por el modelo, mostrando qué tan bien se ajusta el modelo a los datos reales. Su valor varía entre 0 y 1, donde 0 indica que el modelo no explica nada de la variabilidad de los datos y 1 indica un ajuste perfecto (overfitting).

Se define como:

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2},$$

donde  $\bar{y}$  representa el promedio de los valores reales.

### 3.5.3. Error Porcentual Absoluto Medio (MAPE)

Es el cálculo del error promedio en términos porcentuales, usado para medir el error relativo como un porcentaje de la variable objetivo. Es útil cuando se trabaja con datos donde las escalas de las variables son diferentes y facilita la comunicación de la calidad de las predicciones a personas no técnicas.

Definido:

$$MAPE = \frac{\sum \frac{(y_i - \hat{y}_i)}{y_i}}{n} \times 100\%.$$

## 3.6. Grado de ajuste del modelo

Comprender el grado de ajuste del modelo es importante para entender sus capacidades predictivas y garantizar su capacidad de generalización. Observamos el error de predicción en el entrenamiento y evaluación para determinar si el modelo está subajustando o sobreajustándose a los datos de entrenamiento.

### 3.6.1. Subajuste

El underfitting (subajuste) ocurre cuando un modelo no se ajusta lo suficientemente bien a los datos de entrenamiento y, como resultado, no logra capturar las relaciones subyacentes en los datos. Uno de los signos comunes de underfitting es un rendimiento deficiente tanto en el conjunto de entrenamiento como en el conjunto de prueba, ya que el modelo no logra aprender los patrones importantes presentes en los datos y, por lo tanto, no generaliza bien a nuevas instancias. Entre las posibles soluciones a este problema podemos encontrar la utilización de modelos más complejos y el correcto ajuste de hiperparámetros.

### 3.6.2. Sobreajuste

El overfitting (sobreajuste) es un fenómeno que ocurre cuando un modelo se ajusta demasiado bien a los datos de entrenamiento, aprendiendo los detalles y el ruido de los datos sin capturar las verdaderas relaciones subyacentes, perdiendo así su capacidad de generalizar bien a datos nuevos o no vistos. Entre los signos comunes de overfitting se incluye el rendimiento excelente en el conjunto de entrenamiento acompañado de un rendimiento pobre en datos de prueba. Es importante encontrar un equilibrio entre la capacidad del modelo para aprender patrones complejos y su capacidad para generalizar a datos nuevos. Una de las estrategias más utilizadas

para solucionar este problema es la de ajustar los hiperparámetros de los modelos mediante validación cruzada, que será la metodología utilizada para optimizar los modelos.

## 4. Datos

Los datos son la materia prima que utilizaremos para extraer información, conocimiento y, como objetivo final, tomar decisiones basadas en evidencia acerca del precio a asignar a los dispositivos.

### 4.1. Recolección

Con la utilización de la API de Búsqueda de Mercado Libre, que permite obtener todos los ítems listados dentro de una categoría, se extrajo la información diaria de las publicaciones activas de los celulares<sup>10</sup> en venta dentro de Capital Federal<sup>11</sup> en el periodo que va desde el 26/12/2022 hasta el 08/08/2023. Sobre el total de publicaciones encontradas para cada modelo se iteró en conjuntos de 50 (límite determinado por la aplicación) guardando cada código de identificación, título, reputación del vendedor, cantidades disponibles, vendidas, envío y tags asociados.

Luego, con la API de Producto (que permite conocer los atributos asociados a cada publicación en base a su código de identificación) se extrajo la información correspondiente al precio, tiempos, condición y especificaciones de cada una de las publicaciones extraídas. Este proceso fue automatizado para correr cada día de la semana, en la Tabla 1 se detallan las variables recolectadas.

Tabla 1. Descripción de los datos extraídos.

Variable	Descripción	Tipo
ID	código de identificación de la publicación dentro de la plataforma.	Texto alfanumérico
Title	título correspondiente a la publicación.	Texto alfanumérico
Seller_Rep	reputación del vendedor dentro de la plataforma.	Catórica
Avail_Q	cantidades disponibles del producto correspondiente a la publicación.	Numérica discreta
Sold_Q	cantidades vendidas del producto correspondiente a la publicación.	Numérica discreta
Free_Shipping	naturaleza del envío de la publicación.	Binaria

<sup>10</sup> código de categoría MLA1055

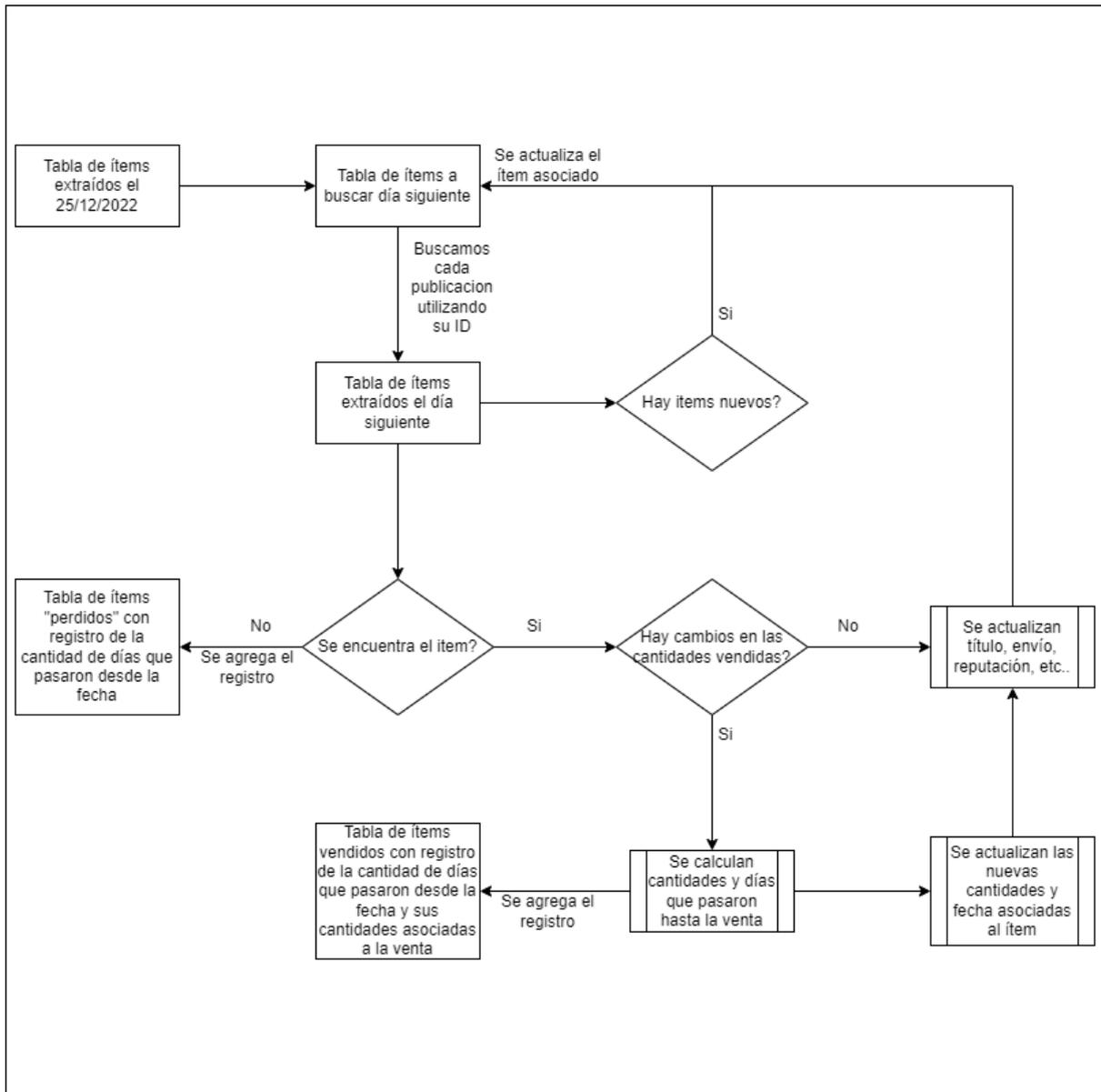
<sup>11</sup> código de ciudad TUxBQ0NBUGZIZG1sYQ

Tags	atributos intrínsecos de la publicación (ejemplo: ["poor_quality_picture", "immediate_payment", "cart_eligible"])	Lista
Model	modelo de celular correspondiente a la publicación	Catógorica
Date	fecha correspondiente a la extracción de los datos	Fecha
Price	precio corriente de venta en pesos argentinos correspondiente a la publicación	Numérica continua
Original_Price	precio original de venta en pesos argentinos correspondiente a la publicación	Numérica continua
Start_Time	fecha de inicio de la publicación dentro de la plataforma	Fecha
Stop_Time	fecha de finalización de la publicación dentro de la plataforma	Fecha
Condition	condición del estado del celular correspondiente a la publicación	Catógorica
Attributes	características del celular correspondientes a la publicación (marca, modelo, memoria interna, RAM, color, procesadores, etc.)	JSON

## 4.2. Diseño

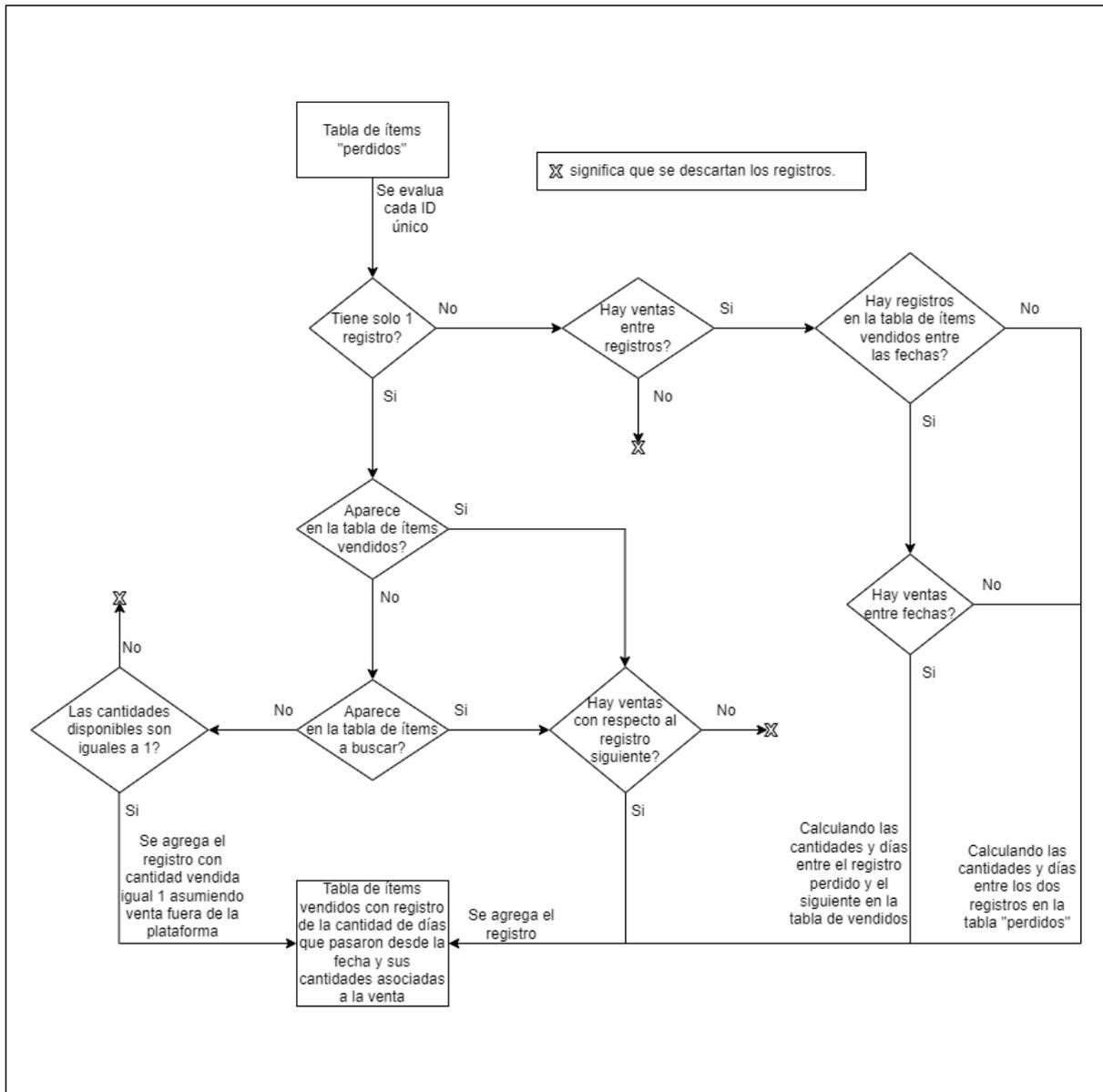
Como se ilustra en la Figura 4, se iteró sobre los datos extraídos guardando los nuevos posteos de celulares y buscando cambios en las cantidades vendidas en los existentes con anterioridad, que fueron almacenados en un dataset "ventas" junto con la cantidad de días que pasaron hasta su realización. Además, se guardó en un dataset "perdidos" aquellos identificadores que no se encontraron al día siguiente. En cada iteración se actualizaron los valores de título, reputación, envío y tags asociados a cada publicación.

Figura 4. Ilustración del diseño del dataset con base a la extracción de la API de Búsqueda.



En la Figura 5 se detalla el preprocesamiento realizado sobre las publicaciones en el dataset “perdidos”, asumiendo como vendidos aquellos ítems que no volvieron a aparecer en la plataforma (y cuyas cantidades disponibles fueran iguales a 1). Las observaciones que tuvieron más apariciones fueron evaluadas de la misma forma que en el proceso anterior en busca de guardar sus ventas (si correspondieran).

Figura 5. Ilustración de la integración de las publicaciones pérdidas dentro del dataset.



Así, contamos finalmente con:

- 10.170 registros de ventas a partir de cambios en las unidades vendidas (que suman 84.344 productos).
- 40.550 registros de ventas por pérdida de la publicación dentro de la plataforma (40.550 productos).

Luego, se agregó la información obtenida acerca de la fecha e identificación de las ventas registradas. En aquellos productos para los cuales no tenemos información en la misma fecha, se buscó la siguiente primera información disponible y, en caso de que no estuviera, la última.

Se encontraron 1.807 casos que no cuentan con información extraída de producto y 13.529 cuya fecha de venta es mayor que la de finalización de la publicación, estos fueron eliminados del análisis.

Para el resto de los registros, se crearon las variables:

- “days\_to\_sale”: Indica la cantidad de días hasta que se produjo la venta.
- “amount\_sold”: Muestra la cantidad de celulares que se vendieron.
- “condition\_of\_sale”: Indica si el celular fue vendido, o si desapareció de la plataforma (y asumimos que fue vendido por afuera).
- “day\_sold”: Indica la fecha en la que se vendió (o desapareció) el celular.

Se transformó la columna “Original\_Price” (que adquiere un valor distinto de nulo solamente cuando el producto presenta un descuento en el precio) para que refleje el porcentaje de descuento que tiene el producto sobre su precio original al momento de recolectar la información, la renombramos como “discount” y la definimos como:

$$discount = 1 - \frac{price}{original\_price} \text{ if } original\_price \neq NaN \text{ else } 0.$$

### 4.3. Análisis exploratorio de los datos recolectados

En la Figura 6 se detalla la cantidad de extracciones realizadas con cada API durante cada mes del año. En el mes de Agosto los días de extracción de los datos es igual con ambas aplicaciones ya que la de producto se utiliza sobre la información extraída en la de búsqueda y, a diferencia de otros meses, su proceso no falló en ninguna ocasión. En la Figura 7 podemos ver un detalle más granular de las extracciones donde se detalla, para cada día del periodo, la aplicación corrida.

Figura 6. Días del mes con extracción de los datos.

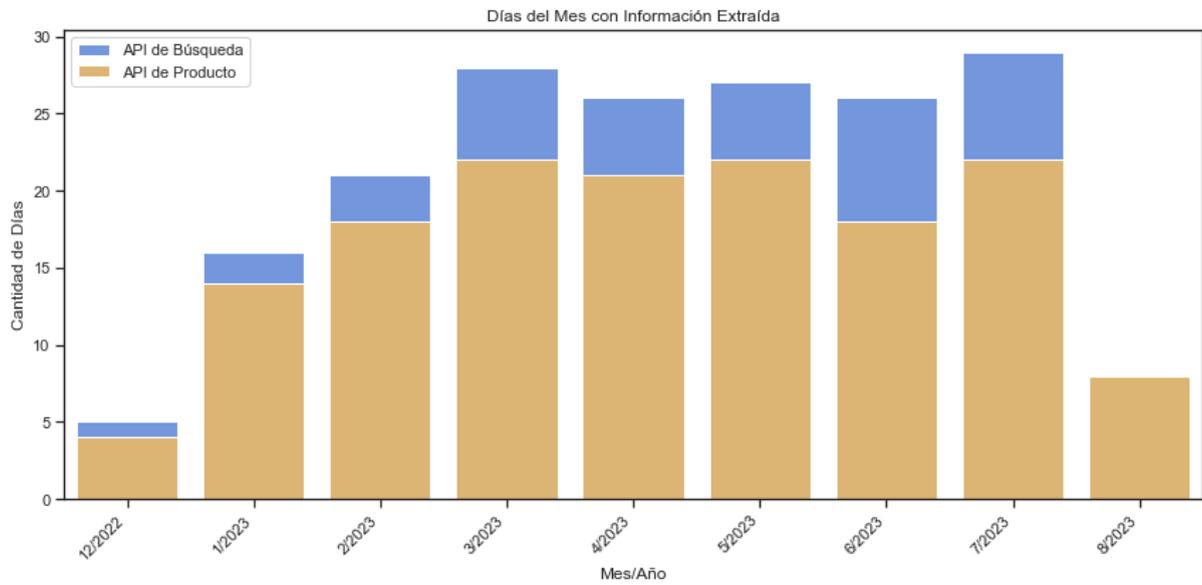
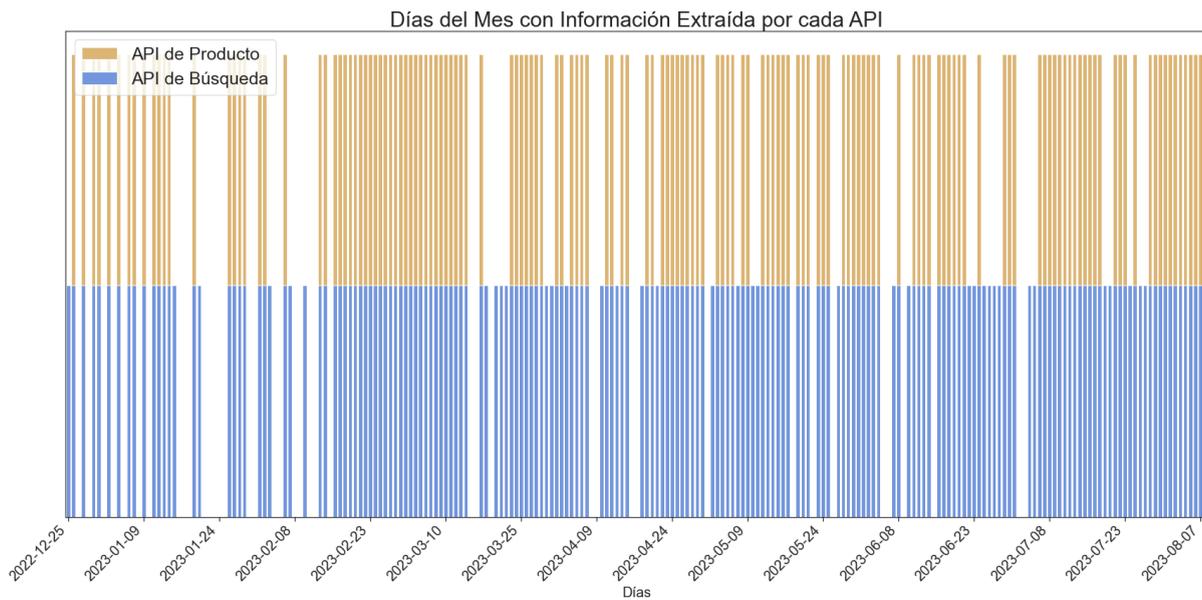


Figura 7. Días del periodo con extracción de los datos.



Como se puede observar, existen fechas en las cuales ambas extracciones no lograron completarse. Por esta razón, y a costa de perder observaciones, se decidió filtrar aquellas ventas marcadas los días siguientes a una falla en la extracción de la API de búsqueda, haciendo que las observaciones en el dataset pasen de 35.384 a 25.684.

A continuación, en las tablas 2 y 3, se analizan las estadísticas descriptivas de las variables.

Tabla 2. Estadísticas descriptivas de variables numéricas.

	Free_Shipping	days_to_sale	amount_sold	condition_of_sale	condition	discount	price
count	25.684,00	25.684,00	25.684,00	25.684,00	25.684,00	25.684,00	25.684,00
mean	0,65	25,39	3,32	0,30	0,41	0,01	253.857,64
std	0,48	37,23	12,62	0,46	0,49	0,04	12.127.421,93
min	0,00	1,00	1,00	0,00	0,00	0,00	250,00
25%	0,00	3,00	1,00	0,00	0,00	0,00	45.999,00
50%	1,00	11,00	1,00	0,00	0,00	0,00	89.499,50
75%	1,00	32,00	1,00	1,00	1,00	0,00	184.999,00
max	1,00	226,00	250,00	1,00	1,00	0,59	1.899.998.100,00

Tabla 3. Estadísticas descriptivas de variables categóricas.

	ID	Model	Title	Seller_Rep	Tags	attributes	day_sold
count	25.684,00	25.684,00	25.684,00	8.742,00	25.684,00	25.684,00	25.684,00
unique	21.794,00	947,00	15.435,00	3,00	1.382,00	14.118,00	159,00
top	MLA1360993070	iPhone 11	iPhone 8 Plus	platinum	['poor_quality_picture', 'immediate_payment', ...	{'id': 'BATTERY_CAPACITY', 'name': 'Capacidad...	14/6/2023
freq	9,00	1.296,00	66,00	7.002,00	7.011,00	218,00	437,00
first	NaN	NaN	NaN	NaN	NaN	NaN	26/12/2022
last	NaN	NaN	NaN	NaN	NaN	NaN	8/8/2023

En base a estas tablas, se concluye que:

- “free\_shipping”: El promedio de envío gratis se encuentra en 0,65, lo que implica que un 65% de las publicaciones presentan el envío gratis del producto.
- “days\_to\_sale”: Los días que pasan hasta que se realizan las ventas van desde el 1 al 226, con un 75% de los casos siendo menores a un mes aproximadamente (32), implicando una considerable demanda de celulares dentro de la plataforma.
- “amount\_sold”: A partir de las observaciones encontramos que en la mayoría de los casos hay una única unidad vendida, pero esta puede alcanzar valores más altos, llegando a extremos de 250 dispositivos.
- “condition\_of\_sale”: Con un promedio de 0,30 y recordando el supuesto de que las publicaciones que desaparecen son vendidas fuera de la plataforma, vemos que el 70% de las ventas fueron realizadas fuera de Mercado Libre.
- “condition”: Catalogando a los celulares usados con un valor de 0, el promedio de 0,41 implica que la mayoría de los celulares vendidos (59%) corresponden a celulares nuevos o reacondicionados.
- “discount”: Vemos que, en general, las publicaciones de celulares dentro del ecosistema no presentan ningún descuento, pero este puede llegar a alcanzar valores de hasta un 59%.
- “price”: El precio de los celulares va desde los \$250 hasta extremos de \$1.899.998.100, con un promedio de \$253.857,64 y un desvío estándar considerablemente alto. Este tipo de valores fue identificado como atípico (ver Sección 4.4 de preprocesamiento, en este también se realiza el ajuste por inflación relevante debido al periodo utilizado en el estudio).
- “ID”: Encontramos publicaciones que presentan más de una venta a lo largo del periodo, llegando hasta un máximo de 9.
- “Model”: En los datos vemos que existen 947 modelos únicos y el más frecuente aparece publicado 1.296 veces.
- “Title”: Vemos 15.435 títulos únicos en las publicaciones, con “iPhone 8 Plus” siendo el más frecuente ( aparece en 66 ocasiones).

- “Seller\_Rep”: Con 3 niveles de reputación únicos que pueden alcanzar los vendedores en la plataforma para ser MercadoLíder<sup>12</sup>, vemos que el más frecuente es el de nivel “platino” con un 80% de ocurrencia, vemos también que predominan los valores nulos, que son casi el 66% de las observaciones.
- “Tags”: Tenemos 1.382 combinaciones únicas de tags en las publicaciones, llegando a repetirse 7.011 veces la más común de estas.
- “attributes”: Con 14.118 sets de atributos asociados a los posts, vemos que solo se repiten 218, lo que implicaría que publicaciones de un mismo modelo de celular tienen distintos valores en este campo.
- “day\_sold”: En el periodo comprendido desde el 26/12/2022 hasta el 8/8/2023, vemos 159 días únicos en los cuales hubo al menos una venta de un celular, con el 14/6/2023 como la fecha donde más ventas presentaron distintas publicaciones (437).

#### 4.3.1. Precio de las publicaciones en el tiempo

Resulta relevante entender de qué manera evolucionaron los precios de los dispositivos a lo largo del periodo analizado ya que Argentina, desde 1960, ha sufrido una inflación por encima de la mediana mundial, esto exceptuando el periodo del régimen de convertibilidad (1994-2001), donde la inflación del país estuvo por debajo de este valor, llegando a estar entre los 10 países con menor inflación en el mundo los años 1996 y 1997 (Ocampo 2017).

La inflación se produce cuando el aumento de los bienes y servicios, que en nuestro caso son los celulares dentro de la plataforma, aumentan de manera sostenida. Cuando hay inflación, significa que el poder adquisitivo del peso argentino disminuye, haciendo que se requiera una mayor cantidad de dinero para comprar el mismo celular. Por lo que es necesario realizar un ajuste de los precios por inflación y este, generalmente, se realiza utilizando el índice de precios al consumidor, que mide los cambios en los precios de los bienes y servicios que compran los hogares. En Argentina, este índice es medido por un organismo público desconcentrado de carácter técnico, que pertenece a la órbita del Ministerio de Economía de la Nación, que dirige todas las actividades estadísticas oficiales que se realizan en la República Argentina, denominado Instituto Nacional de Estadística y Censos (INDEC)<sup>13</sup>.

<sup>12</sup> [MercadoLíder en Mercado Libre, 2023](#)

<sup>13</sup> [INDEC. 2023](#)

En la Tabla 4 se detalla un desglose mensual de este índice para los meses completos del periodo analizado en este trabajo. Vemos que la variación acumulada durante estos 9 meses alcanza un valor del 89,29%.

$$\text{Inflación Acumulada} = (1 + \pi_1) \times \dots \times (1 + \pi_n) - 1,$$

definiendo  $\pi_1, \dots, \pi_n$  como la inflación de cada mes considerado en el periodo.

Tabla 4. Variaciones mensuales del nivel general del IPC, total nacional.

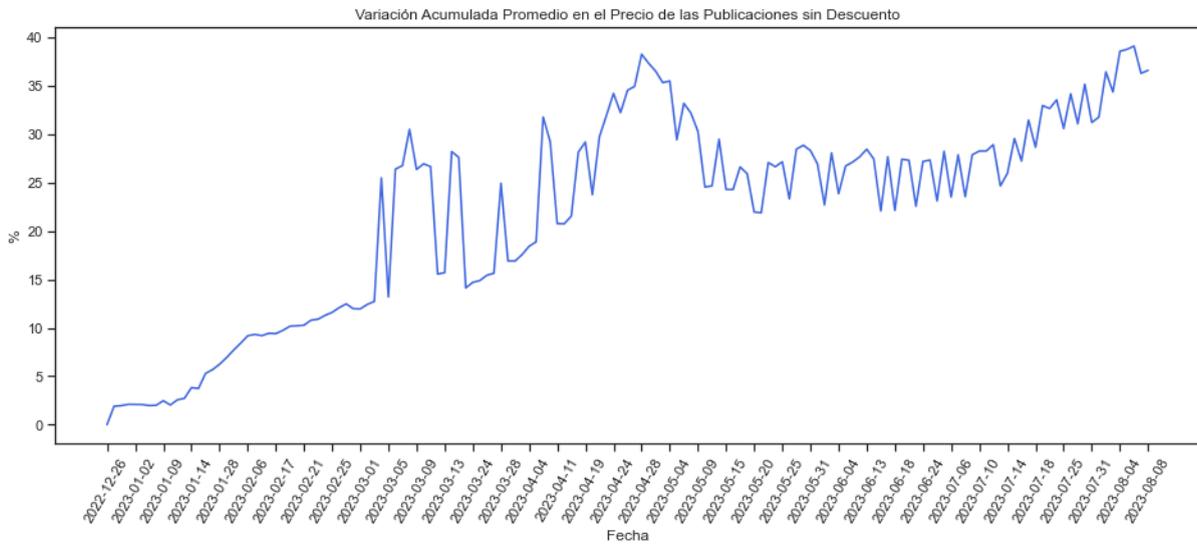
Mes del año 2023	Variación mensual
Diciembre	5,1%
Enero	6%
Febrero	6,6%
Marzo	7,7%
Abril	8,4%
Mayo	7,8%
Junio	6%
Julio	6,3%
Agosto	12,4%

Fuente: [Índice de Precios al Consumidor \(IPC\). Cobertura nacional. Octubre de 2023](#)

Cabe destacar que en este análisis la temporalidad de los datos no corresponde a la mensual obtenida en el IPC, sino más bien a la diaria. Además de esto, el periodo analizado en este trabajo no incluye los meses completos de diciembre y agosto. En la sección de preprocesamiento se distribuyen los valores a ajustar en cada día en base a los valores mensuales correspondientes (ver Sección 4.4.2).

En la Figura 8 podemos observar cuál fue la variación acumulada promedio en el precio de las 4979 publicaciones que presentaron ventas durante su existencia dentro de la plataforma.

Figura 8. Variación acumulada promedio en el precio de las publicaciones en la plataforma.



Exploramos un detalle más granular en la Tabla 5, donde podemos entender cuántas remarcaciones en el precio tuvieron las publicaciones, y cuál fue el incremento acumulado mientras estuvieron dentro de la plataforma (sin filtrar las publicaciones que registraron por lo menos una venta durante su existencia en la plataforma).

Tabla 5. Estadísticas descriptivas de la variación del precio en el tiempo.

	times_changed	accumulated_variation	total_days
count	58.711,00	58.711,00	58.711,00
mean	1,03	51,02	27,62
std	2,23	2.275,15	30,70
min	0,00	-99,92	1,00
25%	0,00	0,00	7,00
50%	0,00	0,00	18,00
75%	1,00	0,00	38,00
max	72,00	366.738,97	148,00

En base a esto, se concluye que:

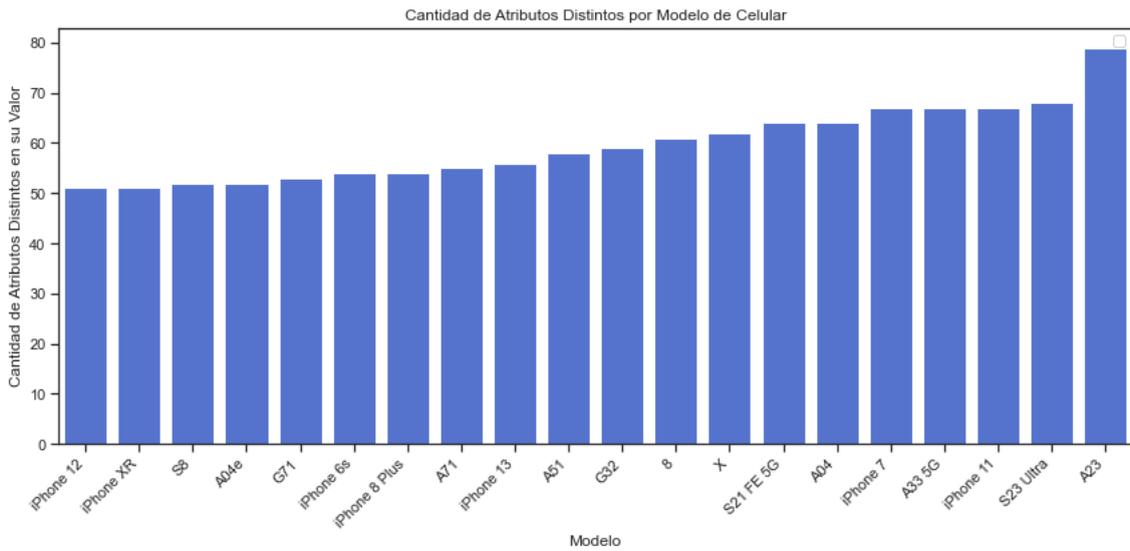
- “times\_changed”: El promedio de veces en las que un vendedor cambia el precio de su celular es cercano a 1, con un máximo observado de 72 y un desvío estándar de 2,23, pero más del 50% de las publicaciones no percibieron cambios en el precio durante su existencia dentro de la plataforma.
- “accumulated\_variation”: En promedio, las publicaciones sufrieron un incremento del 51% dentro de la plataforma, con un desvío estándar de 2275%, el mínimo cambio en el precio observado fue una depreciación del 99,92%, mientras que el máximo se encontró en un aumento acumulado de 366.738,97% en el precio publicado. En la Sección 4.4 de preprocesamiento se evalúan estos valores atípicos para que sean removidos de la información.
- “total\_days”: En promedio, las publicaciones existieron durante 27,62 extracciones realizadas con la API de Producto con un desvío estándar de 30,69, yendo desde los mínimos de 1 día hasta los 148 totales.

#### 4.3.2. Atributos de los celulares

En esta sección analizamos el detalle de la columna “attributes” extraída de Mercado Libre, que contiene las características correspondientes a los celulares publicados. Si transformamos cada uno de los ítems asociados a los dispositivos de los posteos en columnas, encontramos 223 variables a agregar en el dataset, pero como vimos, algunos dispositivos tendrían distinta cantidad de atributos dependiendo de lo que haya cargado el vendedor en la plataforma. Los modelos con distinta cantidad de atributos asociados dentro de la plataforma suman 231.

Si los valores de los atributos asociados fueran estandarizados por la plataforma para cada modelo, podríamos expandir los valores de los que tienen características faltantes utilizando las encontradas en otras publicaciones. En la Figura 9 analizamos los valores distintos asignados a cada modelo de celular en los posteos, donde encontramos 577 casos en los que al menos uno tiene dos o más valores distintos en un mismo atributo, llegando a encontrar modelos con 79 características asociadas distintas.

Figura 9. Últimos 20 modelos de celular con atributos distintos en la plataforma.

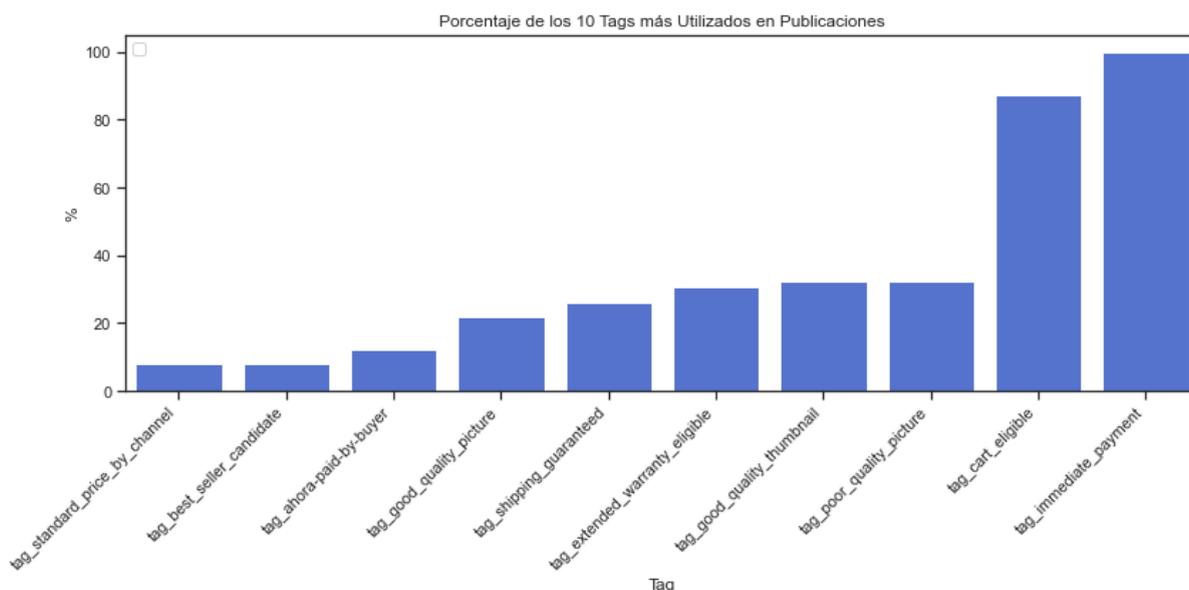


Entre los atributos extraídos, encontramos valores como la marca y el modelo, que también presentan valores distintos en las publicaciones de la plataforma de un mismo “modelo” de celular. Frente a esto, se decidió utilizar la información recolectada de GSMArena (detallado en la Sección 4.3.5) ya que tiene los atributos estandarizados para los distintos modelos, sin embargo, se guardan los atributos de identificación y valor multiple “BRAND”, “MODEL”, “INTERNAL\_MEMORY\_num”, “RAM\_num”, “ITEM\_CONDITION” y “ACCESSORIES\_INCLUDED”.

### 4.3.3. Tags asociados a las publicaciones

En la Figura 10 analizamos los tags asociados dentro de Mercado Libre a las publicaciones, asignando un valor de 1 cuando existe y 0 cuando no. Encontramos que solamente 8 tags aparecen en más del 10% de las publicaciones, con “immediate\_payment” llegando al 100% de los casos.

Figura 10. Utilización de los tags de las publicaciones de la plataforma.



#### 4.3.4. Títulos de las publicaciones

Dentro de los títulos asociados a las publicaciones podemos encontrar información con la cual asignar los atributos faltantes, o corregir los incorrectos. Para esto se separaron los diferentes títulos únicos que aparecieron durante el periodo con el fin de obtener la información correspondiente a los dispositivos utilizando un modelo de lenguaje de la familia PaLM 2, de Google AI<sup>14</sup>.

Del abanico de modelos pre entrenados para elegir dentro de PaLM 2, se utilizó “text-bison@001” con los parámetros estándar (cuyo valor se encuentra en paréntesis al final de la descripción de cada uno):

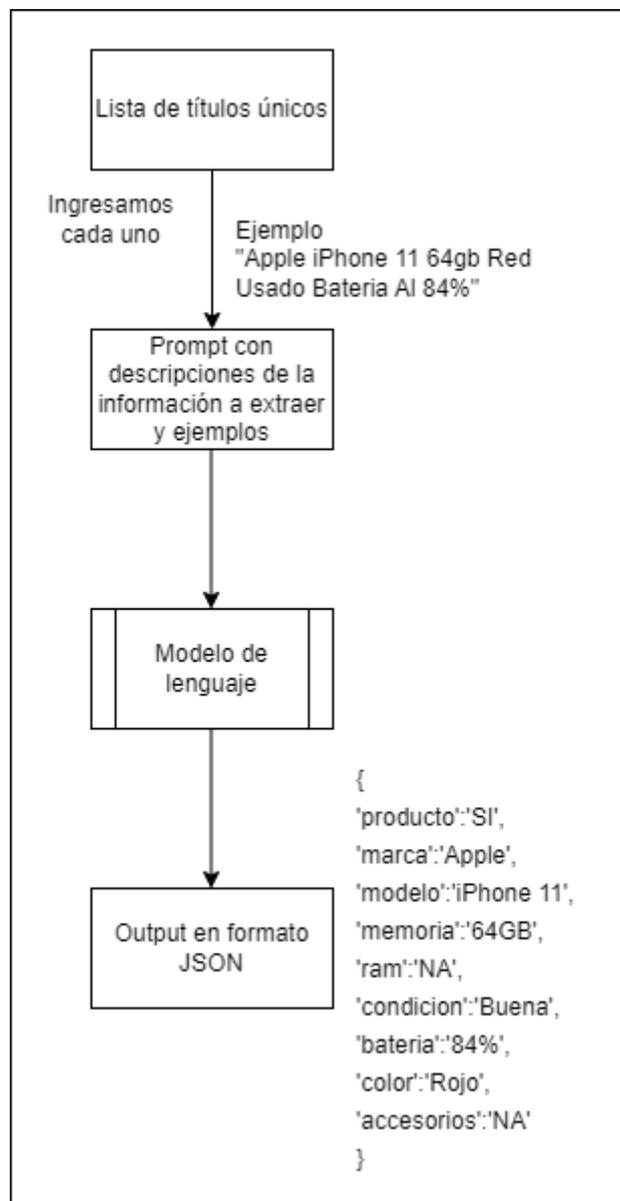
- temperature: Incrementa los pesos que se le da a los otros posibles tokens (después del primero) a la hora de generar una respuesta, cuando buscamos resultados determinísticos, lo recomendable es utilizar un valor bajo así siempre elige el primer token siguiente (0,2).
- top\_p: Determina el threshold de probabilidad que tienen que acumular los próximos posibles tokens para que, una vez alcanzado, deje de considerar nuevos tokens para ser elegidos por el modelo (0,8).
- max\_output\_tokens: Determina el límite de tokens que puede alcanzar la respuesta generada por el modelo (1.024).

<sup>14</sup> [Google AI. 2023](#)

- top\_k: Determina el límite de tokens a considerar para ser seleccionados por el modelo en cada instancia (40).

Los modelos de lenguaje requieren que el usuario ingrese una consulta, denominada “prompt”, para generar una respuesta. El modelo procesa este input mediante la tokenización, que consiste en dividir el texto en unidades más pequeñas, o tokens (que puede ser tan corto como un carácter o tan largo como una palabra completa), para que el modelo pueda procesar y entender la información de manera más efectiva. La Figura 11 presenta una ilustración de este proceso, con ejemplo de título y respuesta del modelo.

Figura 11. Ilustración de la extracción de características de los títulos.



En este proceso, se incluyeron en el contexto de la prompt siete ejemplos de títulos junto con su respuesta en formato JSON para que el modelo aprendiera a generalizar y responder en el

formato correcto a la extracción de la información solicitada. Se presenta un ejemplo en la Tabla 6 con las primeras cinco observaciones del título y su información extraída y debajo, en la Tabla 7 se presentan sus estadísticas descriptivas.

Tabla 6. Vista de títulos e información extraída sobre los productos.

Title	producto	marca	modelo	memoria	ram	condicion	bateria	color	accesorios
iPhone 13 Pro Max 512gb 6gb Ram Grafito Sellad...	SI	Apple	iPhone 13 Pro Max	512GB	6GB	Nuevo	100%	Grafito	NaN
iPhone Lote Repuesto	NO	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
iPhone 11 Pro 256 Gb 81% De Batería Usado	SI	Apple	iPhone 11 Pro	256GB	NaN	Usado	81%	NaN	NaN
Apple iPhone 11 64gb Negro Cargador Cable Fund...	SI	Apple	iPhone 11	64GB	NaN	NaN	NaN	Negro	Cargador, Cable, Funda, Glass
Celular Samsung Galaxy A13 Octacore 128gb 4gb ...	SI	Samsung	Galaxy A13	128GB	4GB	NaN	NaN	Negro	NaN

Tabla 7. Estadísticas descriptivas de variables extraídas.

	producto	marca	modelo	memoria	ram	condicion	bateria	color	accesorios
count	15.426	14.281	14.269	11.861	4.789	5.629	2.562	8.430	2.308
unique	2	62	1.298	65	30	32	62	634	558
top	SI	Apple	iPhone 11	128GB	4GB	Excelente	1	Negro	Cargador
freq	14.153	6.444	797	4.303	1.672	2.941	897	1.797	212

En base a la Tabla 7, concluimos que, de los 15.435 títulos únicos:

- “producto”: Vemos que el modelo no pudo identificar si el título correspondiente a la publicación correspondía a un producto funcional en 9 ocasiones, además, en 1.273 observaciones vemos que el valor asignado a producto es “NO”.
- “marca”: Se identificaron 62 marcas únicas dentro de los títulos, siendo “Apple” la más frecuente con un aproximado de 45% de presencia, en 1.166 casos no se logró identificar ninguna marca dentro del título.
- “modelo”: Se encontraron 1.298 modelos únicos en los títulos, 351 más de los que teníamos asignados dentro de la plataforma, donde sigue prevaleciendo el “iPhone 11” en su mayoría.

- “memoria”: En lo que respecta a la memoria interna del dispositivo, vemos que aparecen 65 valores únicos dentro de los encontrados, siendo el de 128 gigabytes el que aparece en la mayoría de los casos, el modelo no pudo extraer esta información de 3.574 títulos.
- “ram”: Vemos que solo se encontró la capacidad de la RAM en 4.789 títulos, donde la más frecuente corresponde a 4 gigabytes, con 30 capacidades únicas dentro de ellos.
- “condicion”: Se extrajeron 32 condiciones de celular distintas de los títulos, siendo “Excelente” la que tiene mayor cantidad de ocurrencias, con aproximadamente 64% de casos donde no se pudo identificar este valor.
- “bateria”: Siendo la segunda categoría con menor cantidad de información extraída de los títulos, vemos que se extrajeron 62 porcentajes de batería distintos dentro de los títulos, cabe destacar que aproximadamente un 77% de estos corresponden a celulares de la marca “Apple”.
- “color”: Existen 634 colores únicos dentro de los títulos, siendo el Negro el que tiene mayor ocurrencia con un aproximado de 21% de presencia en los 8.430 valores extraídos.
- “accesorios”: Información con menor predominancia dentro de los títulos, pero gran variedad de estos con 558 valores únicos (se destaca que, dependiendo del título, los accesorios extraídos pueden corresponder a más de uno, y, con un distinto orden, vemos este número mayor del real que debiéramos encontrar después de realizar una limpieza de la variable), siendo el cargador el que tiene mayor presencia con 212 observaciones.

#### 4.3.5. Atributos estandarizados de GSMarena

Como se describió en la Sección 4.3.2, existen distorsiones en los atributos asociados a los distintos modelos de la plataforma que agregan ruido en el modelo cuando deberían ayudarnos a conseguir mejores resultados en términos de mayor predictibilidad. Por esta razón se decidió reemplazarlos por los obtenidos de GSMarena, portal dedicado a la telefonía móvil con una extensa cantidad de productos y una amplia y más estandarizada base de datos con especificaciones técnicas.

Se creó una tabla que combina las distintas columnas de marca (una obtenida de los atributos asociados y la otra extraída de los títulos) y modelo (la primera obtenida de los atributos, la segunda de los títulos y la tercera en la recolección vía API) con un total de 2.322 combinaciones (la Tabla 8 ofrece una vista de las primeras 5 combinaciones). A estas combinaciones se les

asignó, en los casos donde el producto existe dentro de la página, de forma manual el hipervínculo asociado dentro del portal y sus valores de marca, modelo y línea.

Tabla 8. Vista de tabla con combinación de marcas y modelos del dataset.

BRAND	marca	MODEL	modelo	Model
Apple	Apple	iPhone 8	iPhone 8	iPhone 8
Apple	NaN	iPhone 6s	NaN	iPhone 6s
Apple	Apple	iPhone	iPhone XS	iPhone
Samsung	Samsung	J7	J6	J6
Apple	Samsung	iPhone XR	Galaxy S20	iPhone XR

Luego de tener los vínculos asignados, utilizando la librería de Python Beautiful Soup, se extrajo la información técnica de los celulares (en formato de texto alfanumérico) detallada a continuación en la Tabla 9.

Tabla 9. Especificaciones descritas en GSMarena.

Variable	Descripción
Network	coberturas soportadas por el celular.
Status	disponibilidad actual y año de lanzamiento del celular.
Dimensions	medidas correspondientes al cuerpo del celular (alto, largo y ancho), en milímetros.
Weight	peso del celular, en gramos.
Size	tamaño de pantalla del celular, junto con su ratio correspondiente al cuerpo.
Resolution	píxeles, ratio y densidad en píxeles por pulgada de la pantalla del celular.
CPU	unidad central de procesamiento (más conocido como procesador) del celular.
Card Slot	indicador de habilitación de memoria externa del celular.
Internal Memory	combinaciones posibles de memoria y RAM del celular.

Main Camera	categorización de la cantidad de cámaras principales correspondientes al celular.
Main Camera Specs	especificaciones técnicas de cada una de las cámaras principales del celular.
Selfie Camera	categorización de la cantidad de cámaras frontales del celular.
Selfie Camera Specs	especificaciones técnicas de cada una de las cámaras frontales del celular.
3.5mm Jack	indicador de habilitación de conector estándar de audio en el celular.
NFC	indicador de disponibilidad de tecnología del celular.
Radio	indicador de disponibilidad de tecnología del celular.
Features	sensores que se encuentran integrados al celular.
Battery	tipo y capacidad de la batería correspondiente al celular, en miliamperios-hora.
Battery Life	calificación de resistencia correspondiente a la batería del celular, en horas.

A continuación, se destacan las transformaciones realizadas a los atributos:

- Network: Se transforma en variable numérica que muestra la máxima cobertura que tiene cada celular, ya que cada celular contiene todas las coberturas anteriores a esta.
- Status: Se crean dos variables, donde la primera indica la disponibilidad del celular (si este se encuentra discontinuado o no), y la segunda la fecha de su lanzamiento, que luego se utiliza para calcular la cantidad de días de vida que tiene el modelo de celular desde su lanzamiento hasta la fecha en la que es vendido.
- Dimensions: Se multiplican la altura, ancho y largo del celular para expresar el volumen del celular en milímetros cúbicos.
- Weight: Filtramos los valores para utilizar solo el número expresado en gramos.

- Size: Filtramos los valores para utilizar solo el número correspondiente al ratio pantalla/cuerpo del celular.
- Resolution: Se crean tres variables, en la primera multiplicamos los pixeles de la pantalla para mostrar los pixeles cuadrados, en la segunda se divide el ratio para utilizar los valores precisos de este, y en la tercera dejamos la densidad de pixeles por pulgada.
- CPU: Filtramos únicamente la cantidad de procesadores correspondiente, y luego transformamos el valor en numérico.
- Card Slot: Transformamos en 0 y 1, siendo 1 el celular con disponibilidad de memoria externa.
- Internal Memory: Se crea una tabla que contiene, para cada modelo de celular (a través de su link en la página, usado de identificador), las combinaciones posibles de memoria interna y RAM correspondientes. Llegando a un total de 652 para los modelos seleccionados que se utilizan como determinante en los casos en los que los valores del título y atributos no coinciden en la publicación.
- Main Camera: Se transforma en variable numérica, indicando la cantidad de cámaras principales contenidas.
- Main Camera Specs: Creamos 4 variables, las dos primeras muestran el mínimo y máximo de megapíxeles entre las distintas cámaras, y las siguientes el mínimo y máximo de apertura que tienen.
- Selfie Camera: Se transforma en variable numérica indicando la cantidad de cámaras frontales presentadas.
- Selfie Camera Specs: Al no contar con cámaras potentes en todos los celulares, se decide utilizar únicamente el máximo de megapíxeles que alcanza el celular entre las cámaras frontales contenidas.
- 3.5mm Jack: Transformamos en 0 y 1, siendo 1 el celular con disponibilidad de conector estándar de audio.
- NFC: Transformamos en 0 y 1, siendo 1 el celular con tecnología correspondiente.
- Radio: Transformamos en 0 y 1, siendo 1 el celular con radio integrada.

- Features: Se crean 15 columnas (cada una correspondiente a cada sensor único contenido entre los celulares) binarias que indican la disponibilidad del atributo en el celular.
- Battery: Se crean 2 columnas, la primera indicando si la batería corresponde a una de Li-Po o Li-ion, y la segunda que muestra los miliamperios-hora correspondientes a esta.
- Battery Life: Filtramos únicamente el número de calificación numérico de resistencia obtenido por el dispositivo.

## 4.4. Preprocesamiento

En esta sección se detalla el preprocesamiento aplicado a los datos recolectados de las distintas fuentes, que incluye su integración, el ajuste por inflación aplicado a los precios expresados en pesos argentinos en cada día y un análisis de los registros de celulares disponibles, con eliminación de outliers en el precio ajustado. También, se transformó la variable objetivo para mejorar el poder predictivo de los modelos.

### 4.4.1. Integración de las fuentes de datos

Cuando se integraron los datos extraídos de GSMarena se encontraron 13.563 casos que no contenían información dentro de la página, por lo que se removieron del dataset llevándolo de 25.684 observaciones a 12.121. Luego, se corrigieron los valores de memoria y RAM de los celulares utilizando aquellos que corresponden con las opciones extraídas de la página, asignando valores nulos cuando ninguna de las opciones disponibles coincidiera con estos.

Los valores nulos encontrados en la memoria y la RAM alcanzan las 42 y 953 observaciones respectivamente. Para intentar completar los valores de RAM, se iteró sobre los casos donde la memoria estaba disponible, asignando su valor correspondiente si este podía tener solo una opción en las combinaciones, logrando llenar 920 de 926 observaciones que se encontraban sin su valor. Este proceso se repitió inversamente para rellenar 4 de 15 valores faltantes en la memoria, disminuyendo los faltantes a 38 en el caso de la memoria y 33 en el de la RAM.

Con 200 modelos únicos, en la Tabla 10 se muestra el porcentaje de valores nulos en las columnas que los presentan (sobre un total de 90 columnas):

Tabla 10. Variables con porcentaje de valores nulos.

Variable	Porcentaje de Nulos
accesorios	89,68%
accessories_included	89,34%
bateria	86,87%
seller_rep	75,52%
condicion	69,28%
memory_num	0,31%
memory_unit	0,31%
ram_unit	0,27%
ram_num	0,27%

Debido a la gran cantidad de valores nulos, se descartaron las columnas “accesorios”, “accessories\_included”, “bateria” y “condicion”. Sobre la columna “seller\_rep” se asumió que los valores faltantes son los vendedores sin reputación dentro de la plataforma y se les asignó un 0, “silver”, “gold” y “platinum” recibieron 1, 2 y 3 respectivamente. En la columna “item\_condition” también transformamos “Usado”, “Reacondicionado” y “Nuevo” en 0, 1 y 2.

Para los faltantes en la memoria y ram, se decidió eliminar del dataset las observaciones. Luego, se multiplicaron las unidades de memoria expresadas en terabytes por 1.000, y se eliminaron “memory\_unit” y “ram\_unit” (que contienen un único valor dentro de toda la data).

Se utilizaron las variables “release\_date” y “day\_sold” para calcular la cantidad de días transcurridos desde la fecha de lanzamiento del celular hasta el día registrado de venta, guardados en una nueva variable “age\_in\_days”. Luego, se eliminó del dataset la fecha en la que el dispositivo salió a la venta y la fecha en la que se vendió.

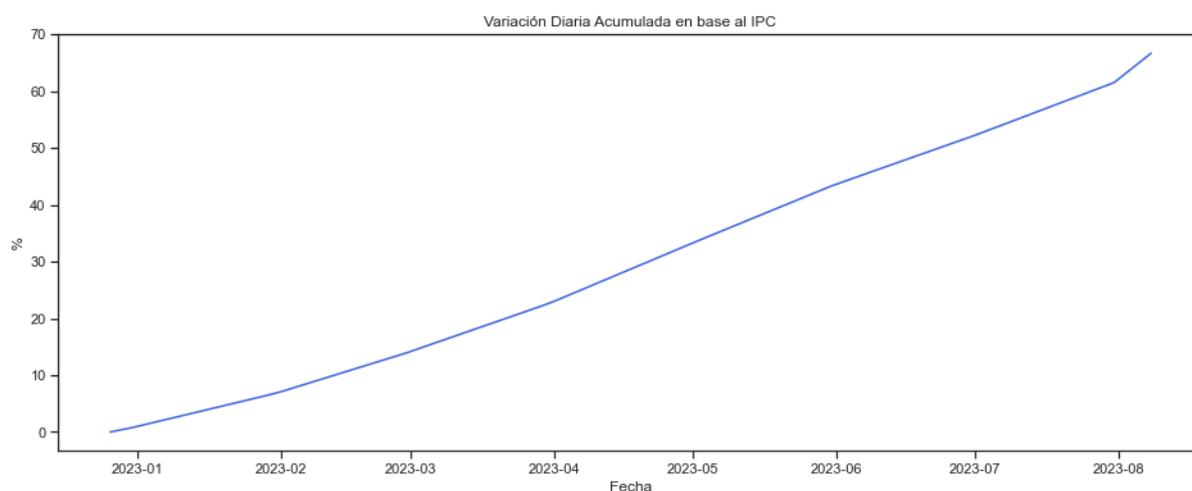
Se dividieron los días registrados hasta la venta por las cantidades vendidas del celular, redondeando al valor entero más cercano los días necesarios para vender un teléfono, creando la variable “time\_to\_sell” y eliminando “amount\_sold” y “days\_to\_sale”. En la sección siguiente se aborda el problema de la nominalidad en el precio de las publicaciones.

#### 4.4.2. Ajuste por inflación de precios de publicaciones

Como adelantamos en la Sección 4.3.1, es necesario realizar un ajuste diario de los precios de las publicaciones. Para esto se creó un dataframe que consiste de las fechas comprendidas desde el 1 de diciembre del 2022 hasta el 31 de agosto del 2023 que incluye en cada fin de mes un índice (con base 1 que se asigna el primero de diciembre) de la inflación acumulada hasta el momento en base al dato del IPC mensual correspondiente.

Luego, se decidió interpolar linealmente los días para los cuales no tenemos un valor asignado (asumiendo que no hay estacionalidad en la inflación dentro del mes), se filtraron las fechas en las que no se extrajo la información de los precios, y se utilizó el índice para calcular la inflación diaria asignada a cada fecha. Se asignó como base 1 el 26 de diciembre del 2022 que fue multiplicado por la inflación correspondiente a cada día para así obtener el índice de la inflación acumulada. En la Figura 12 podemos observar la inflación diaria acumulada obtenida, que alcanza un valor del 66,71% el 8 de agosto del 2023.

Figura 12. Variación diaria acumulada durante el periodo en base al IPC.



Se utilizó la siguiente fórmula para calcular el multiplicador que lleva los precios de cada día a los precios del fin del periodo.

$$Multiplier = \frac{Ind_{8-8-2023}}{Ind_i},$$

donde  $i$  corresponde a cada fecha comprendida en el análisis y el precio de cada publicación es multiplicado por su multiplicador correspondiente.

Después de ajustar los precios, analizamos en la Tabla 11 las estadísticas descriptivas correspondientes, donde encontramos que el 1% de las observaciones (121) no alcanza los

11.675 pesos argentinos del 8 de agosto de 2023, y que 8 registros (0,06%) superan los \$3.589.036, llegando a un valor máximo de 2.849.901.031,90 pesos. Estos valores se filtran manualmente por considerarse atípicos en el dataset.

Tabla 11. Estadísticas descriptivas del precio ajustado.

	count	mean	std	min	1%	50%	95%	max
adj_price	12.077,00	576.314,73	26.359.999,69	259,49	11.675,46	183.826,68	647.304,85	2.849.901.031,90

En la sección siguiente analizamos el precio ajustado de los modelos de celulares disponibles, considerando no solamente el modelo, sino también su memoria, RAM y condición.

#### 4.4.3. Identificación y eliminación de registros problemáticos, separación del dataset en conjuntos de entrenamiento y evaluación

Si combinamos los 196 modelos únicos que tenemos con sus distintas capacidades de memoria y RAM, más los distintos estados de condición en los que aparece dentro de la plataforma, obtenemos 639 identificadores que van desde 1 observación hasta las 530, con un promedio de 18,7 y un desvío estándar de 47,5.

Trabajar con diferentes combinaciones presenta desafíos adicionales ya que celulares de distintas capacidades y en distinto estado pueden tener variaciones importantes en el precio. Con el objetivo de que cada conjunto disponible esté representado, determinamos un límite de observaciones en el grupo equivalente al promedio redondeado, por lo que se filtró todos aquellos grupos con menos de 20 registros, que equivalen a 500 conjuntos que representaban un 19,87% de nuestro dataset.

Se separó la información en conjuntos de entrenamiento y evaluación dejando 85% y 15% de la información respectivamente y, utilizando el precio ajustado, se analizaron los outliers detectados mediante el método de rango intercuartílico, que considera outliers aquellas observaciones que se encuentran 1,5 veces el IQR debajo del primer cuartil, y 1,5 veces el IQR arriba del tercer cuartil.

$$\text{Lower Bound} = Q1 - 1,5 \times IQR,$$

$$\text{Upper Bound} = Q3 + 1,5 \times IQR,$$

donde  $Q1$  es el percentil 25 en el conjunto de datos,  $Q3$  el percentil 75 e  $IQR$  la diferencia entre estos ( $Q3 - Q1$ ).

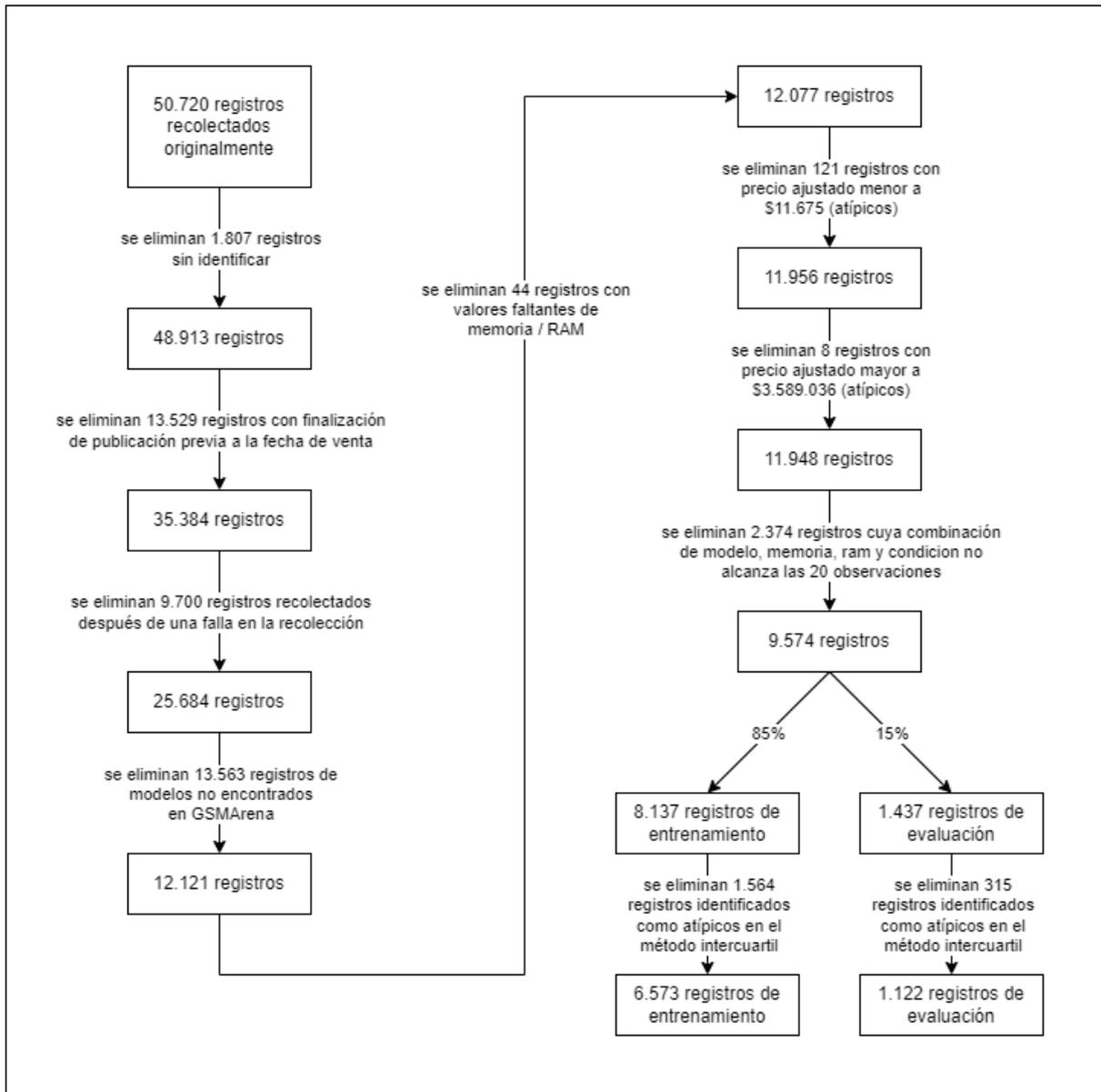
El análisis se realizó sobre cada combinación de modelo, memoria, RAM y condición de celular determinando los valores con los datos de entrenamiento. En la Tabla 12 se destacan las cantidades y porcentajes que representan.

Tabla 12. Análisis de valores atípicos en el precio ajustado.

Variable	#outliers train	%outliers train	#outliers test	%outliers test
adj_price	1.564	19,22%	315	21,92%

Después de eliminar aquellas observaciones catalogadas como outliers nos quedamos con 6.573 registros de entrenamiento para el modelo de predicción y 1.122 para la validación de este (en la Figura 13 se detalla cada eliminación realizada sobre los datos). Se eliminaron las columnas que quedaron con valores únicos en el dataset de entrenamiento ya que no contienen información que le pueda dar poder de predicción al modelo, siendo estas; "tag\_9x\_campaign", "tag\_ahora-12", "tag\_immediate\_payment", "tag\_lightning\_deal", "tag\_only\_html\_description", "tag\_price\_review", "feature\_baroreceptor", "feature\_coulombmeter" y "feature\_color\_spectrum".

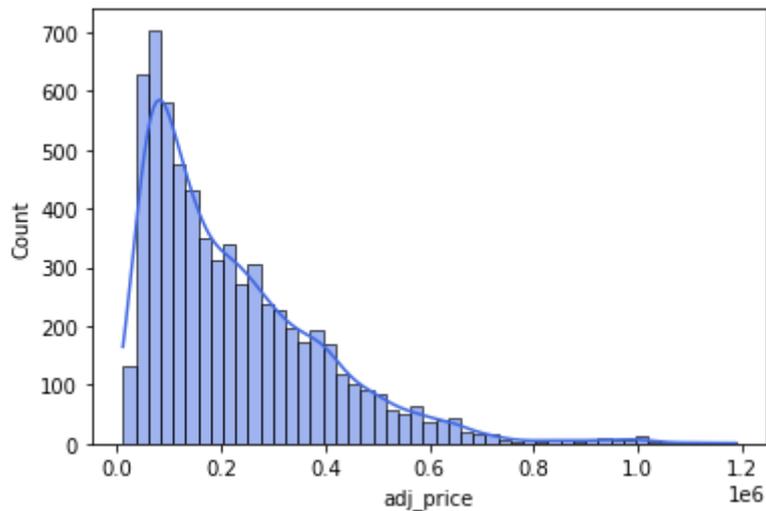
Figura 13. Ilustración con el detalle de la eliminación de registros.



#### 4.4.4. Transformación de la variable objetivo

Si examinamos la distribución de la variable objetivo “adj\_price” en la Figura 14 encontramos la existencia de una asimetría positiva donde la cola del lado derecho es más larga debido a que la mayoría de los valores aparecen en el lado izquierdo.

Figura 14. Histograma de la variable objetivo en los datos de entrenamiento.



La distribución asimétrica es un problema en nuestro contexto ya que:

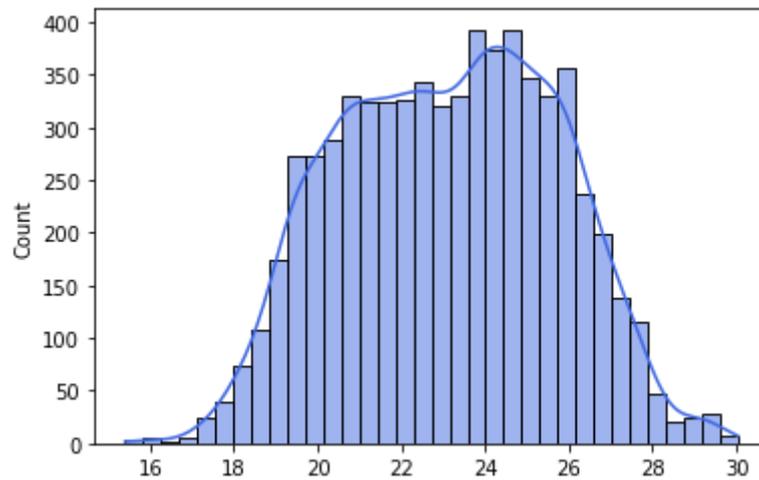
- Muchos de los modelos de regresión, como la regresión lineal, asumen una distribución normal de los errores.
- Los valores atípicos en el lado derecho pueden tener un impacto desproporcionadamente grande en la estimación de los parámetros del modelo cuando éste minimiza los errores cuadráticos.
- Si la mayoría de los valores están en el extremo inferior de la escala y hay pocos valores más altos, el modelo puede tener dificultades para hacer predicciones precisas para los valores más altos.
- Las distribuciones asimétricas positivas a menudo están asociadas con la presencia de valores atípicos en el lado derecho de la distribución. Estos valores atípicos pueden afectar significativamente la calidad del modelo y conducir a predicciones menos confiables.

Hacer que la distribución sea más simétrica puede facilitar la modelización y mejorar la interpretación de los resultados. Entre las transformaciones más comunes para resolver este problema encontramos la logarítmica, raíz cuadrada y box-cox, que incluye la logarítmica y otras transformaciones parametrizadas. Esta última tiene la ventaja de automatizar el tipo de transformación, delegando su elección al modelo.

Evaluando la falta de simetría de las distintas transformaciones se determinó que la más simétrica es la obtenida mediante box-cox, que podemos visualizar en la Figura 15. Luego de

determinar su valor de lambda en los datos de entrenamiento, esta transformación se utilizó para transformar también los datos de validación.

Figura 15. Histograma de la variable objetivo transformada en los datos de entrenamiento.



A la hora de evaluar los resultados obtenidos por los modelos en la sección siguiente se aplicó la inversión de la transformación para que estos se encontraran en la escala original.

## 5. Resultados y análisis de las predicciones

En esta sección utilizaremos los datos definidos anteriormente para entrenar distintos modelos de regresión capaces de predecir el precio de mercado de los dispositivos. Luego, comparamos sus resultados tanto en el set de entrenamiento como en el de evaluación. Las métricas a evaluar en cada caso serán el coeficiente de determinación, raíz del error cuadrático medio y error de porcentaje medio absoluto, incluyendo a su vez el análisis de los residuos y errores porcentuales.

### 5.1. Evaluación de modelos base

Iniciamos esta evaluación sin realizar ninguna selección específica de las variables predictoras ni ajustar los hiperparámetros de los modelos (utilizando aquellos que vienen por defecto). Esto nos permite entender el comportamiento inicial de los modelos e identificar rápidamente las limitaciones que presentan. Además, el modelo elegido en esta instancia funcionará como punto de comparación con los modelos optimizados, permitiendo evaluar si los cambios subsiguientes realmente mejoran el poder predictivo o solo agregan complejidad innecesaria.

En la Tabla 13 se muestran los resultados en los datos de entrenamiento y evaluación para los modelos de Regresión Lineal, Árbol de Decisión, Random Forest y XGBoost.

Tabla 13. Evaluación de los modelos de predicción sin hiperparametrización.

Model	Train R2	Test R2	Train RMSE	Test RMSE	Train MAPE	Test MAPE
Linear Regression	95%	95%	36.819,89	40.682,08	13,19%	13,80%
Decision Tree	100%	95%	4.778,41	41.012,22	0,39%	14,96%
Random Forest	99%	97%	12.289,35	32.709,10	4,38%	12,04%
XGBoost	98%	97%	21.780,68	32.211,75	7,54%	11,95%

Analizamos los resultados de las métricas de cada modelo:

- La Regresión Lineal explica el 95% de la variabilidad del precio tanto en el entrenamiento como en la evaluación y, en promedio, durante el entrenamiento presenta un desvío de 36.819,89 pesos que representa un error porcentual del 13,19%, mientras que en los datos de evaluación el desvío se incrementa hasta los \$40.682,08, aumentando también el error porcentual en 0,61 puntos.
- En el Árbol de Decisión ya podemos notar indicios de overfitting (sobreajuste) ya que, durante el entrenamiento, el modelo logra explicar toda la variabilidad del precio teniendo un desvío promedio de apenas \$4.778,41 que representan un error porcentual promedio de 0,39%, pero estos resultados no son replicados en la evaluación, donde la variabilidad explicada cae 5 puntos porcentuales y, en promedio, el desvío es de 41.012,22 pesos que representan un error porcentual de 14,96%.
- El modelo de Bosque Aleatorio logra explicar un 99% de la variabilidad en entrenamiento con, en promedio, 12.289,35 pesos de desvío y un error porcentual de 4,38%. En la evaluación, la explicación de la variabilidad cae hasta el 97% y, en promedio, el desvío en precios sube hasta los \$32.709,10, con un error porcentual de 12,04%.
- Durante la etapa de entrenamiento, el XGBoost consigue explicar el 98% de la variabilidad en el precio, con un desvío promedio de estos en 21.780,68 pesos y un error porcentual de 7,54%. En la evaluación, la explicación de la variabilidad cae en 1 punto porcentual mientras que, en promedio, el desvío sube hasta los 32.211,75 pesos y el error porcentual alcanza el 11,95%.

A continuación, se visualizan los valores actuales y predichos, residuos y el porcentaje de error de cada uno de los modelos con el fin de obtener conclusiones acerca de estos. La diagonal presente en las Figuras 16 y 17 ilustran los valores que tendría un modelo que predice perfectamente el precio de los dispositivos (como se destaca, una aproximación a esto sucede en el modelo del árbol cuando se comparan las predicciones en los datos de entrenamiento).

Las Figuras 18 y 19 detallan los residuos de las predicciones de cada modelo, las Figuras 20 y 21 el error porcentual presente sobre los datos de entrenamiento y evaluación respectivamente. Los puntos arriba de la línea del cero presente en estos gráficos significan que el modelo asignó un valor menor al observado en el precio del celular mientras que los puntos debajo indican una sobrevaloración del dispositivo.

Figura 16. Valor actual vs. predicho con línea de predicción perfecta en entrenamiento.

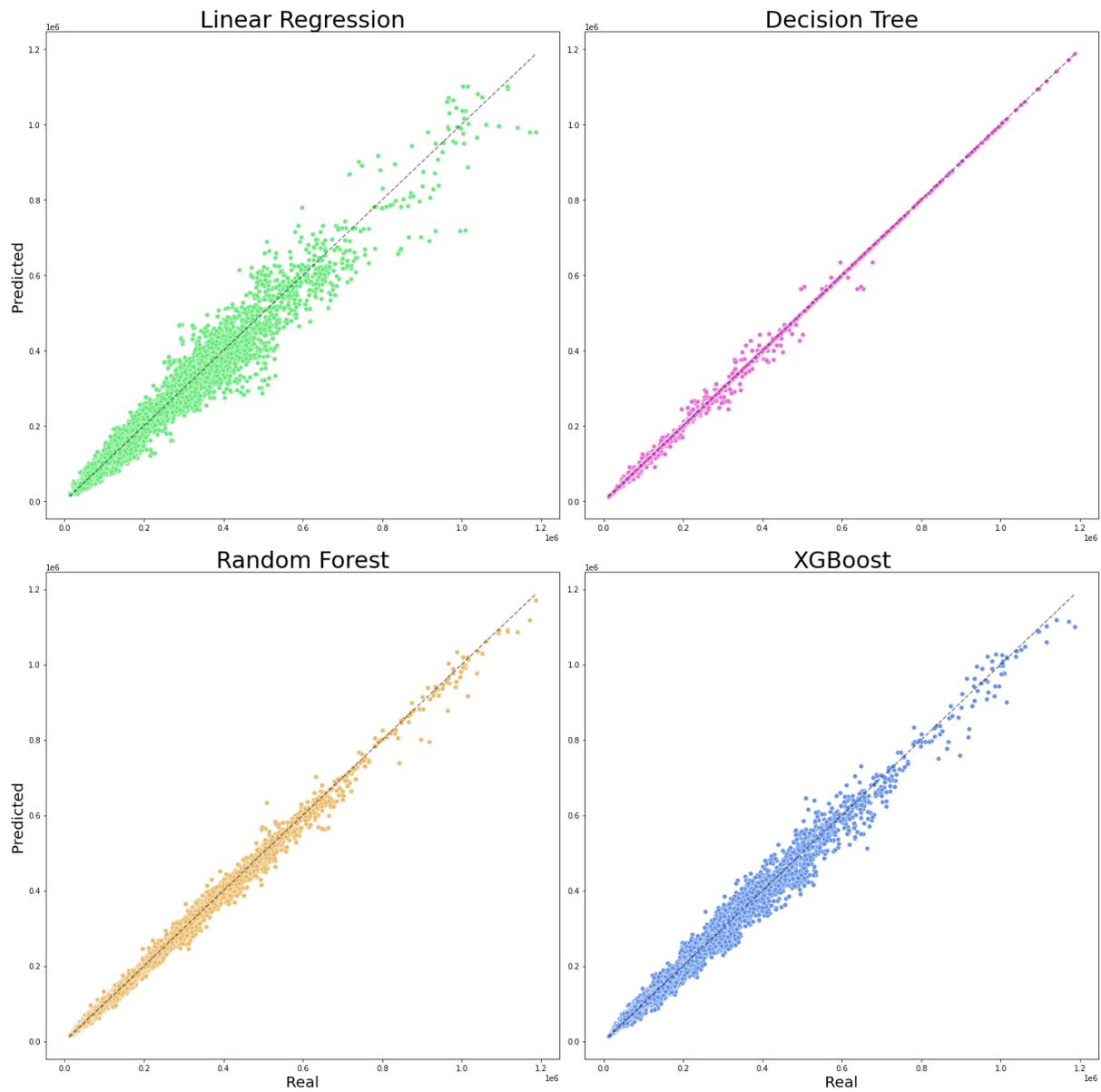


Figura 17. Valor actual vs. predicho con línea de predicción perfecta en evaluación.

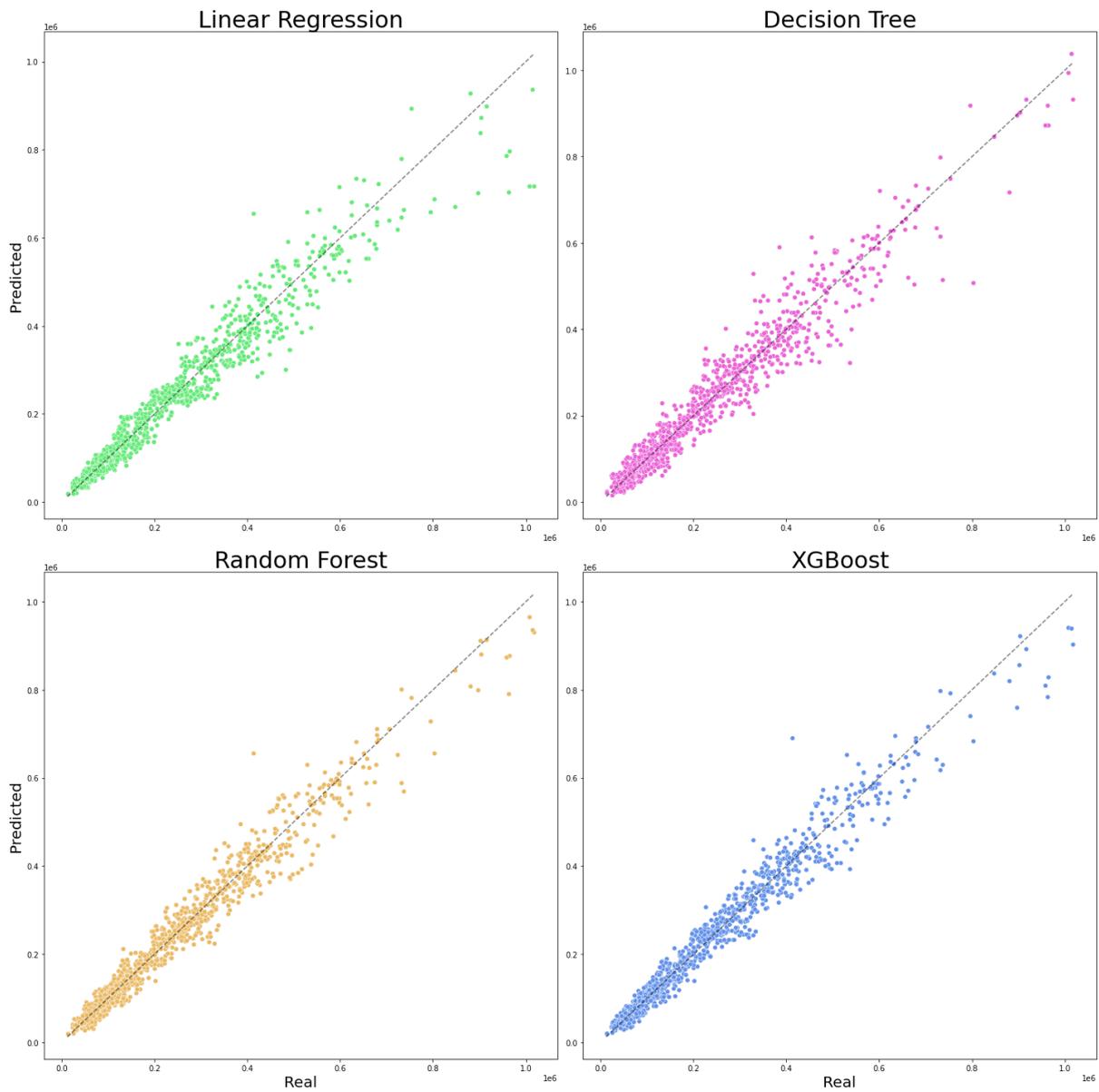


Figura 18. Residuos en entrenamiento.

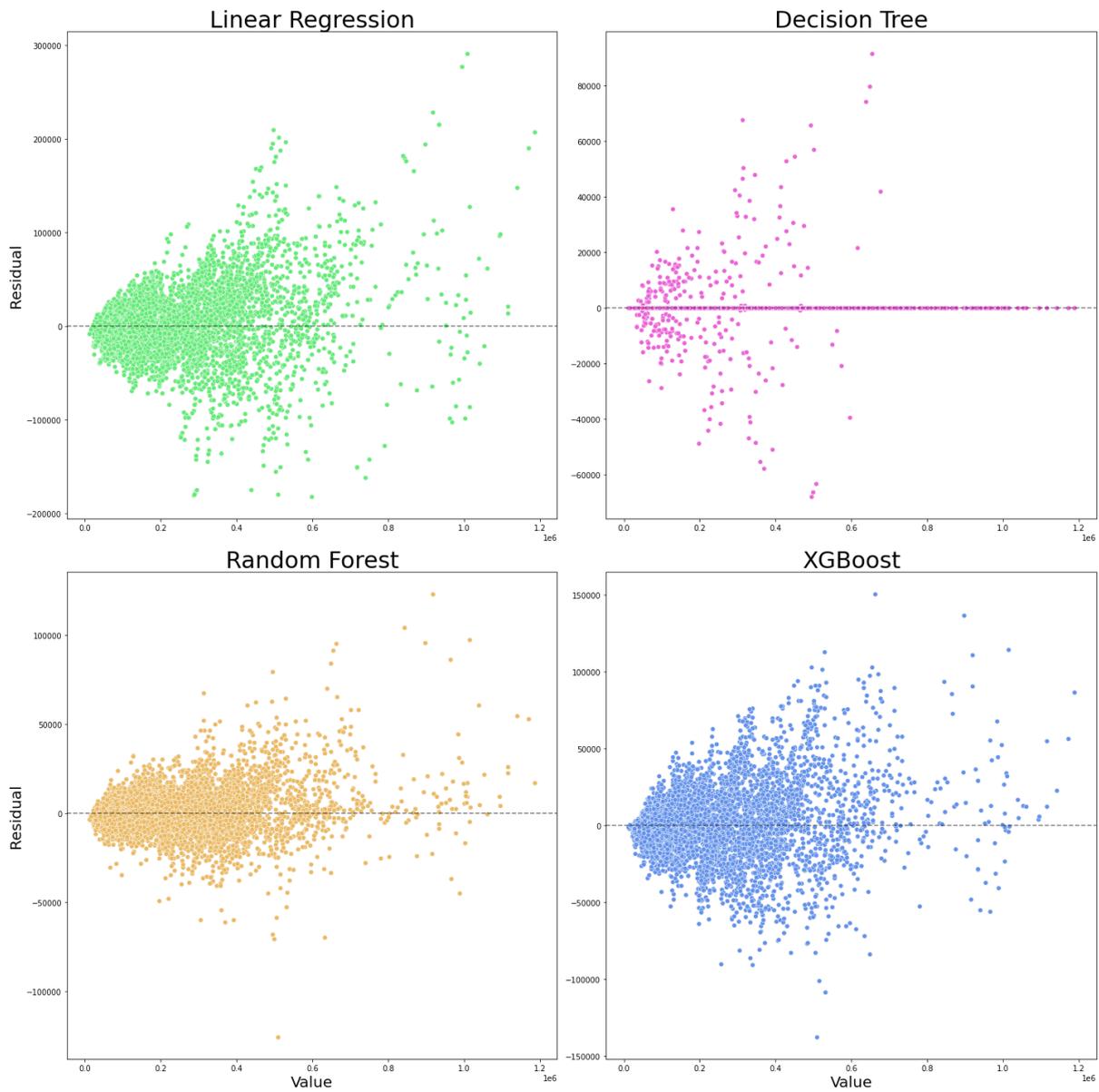


Figura 19. Residuos en evaluación.

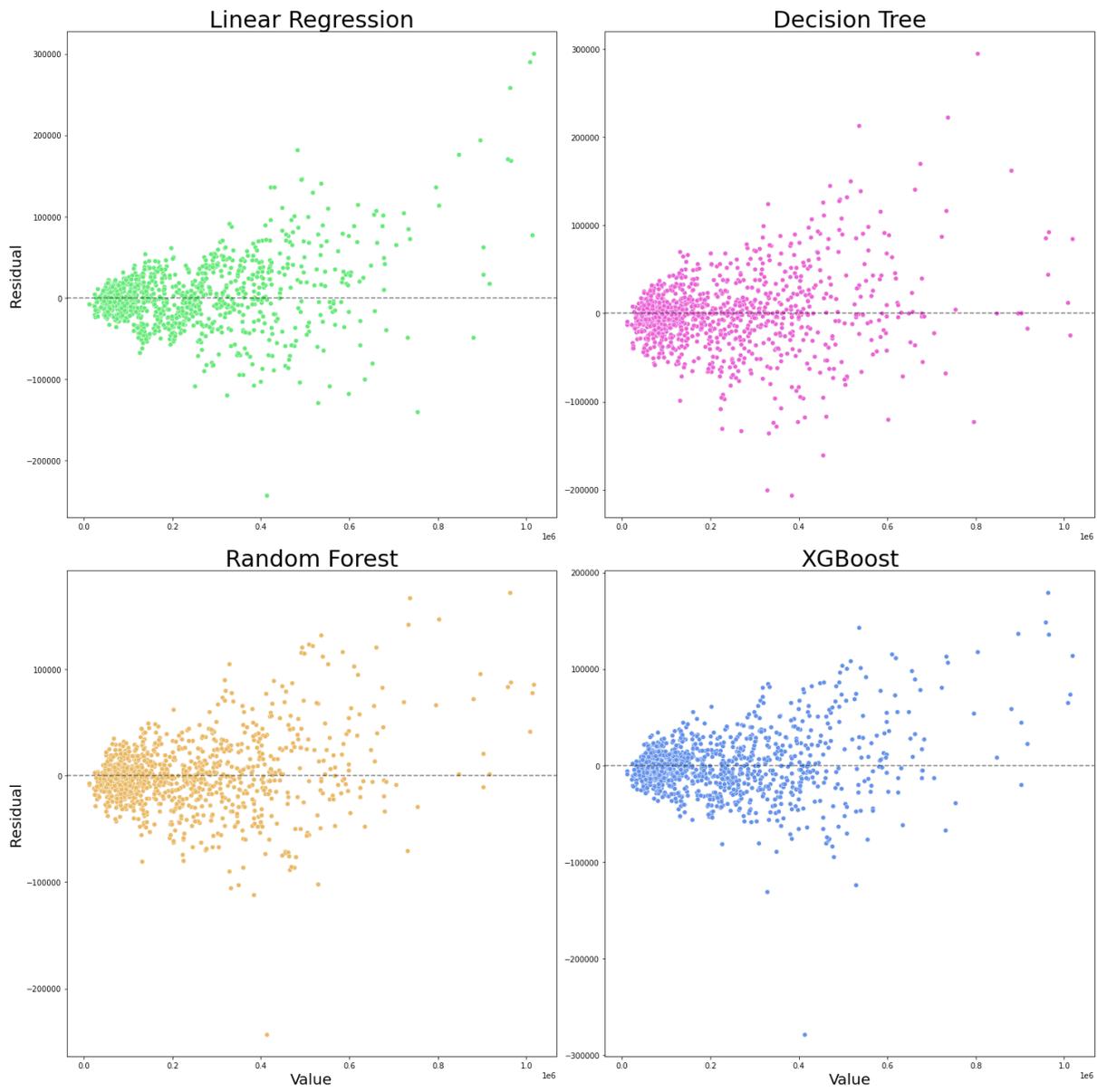


Figura 20. Error porcentual en entrenamiento.

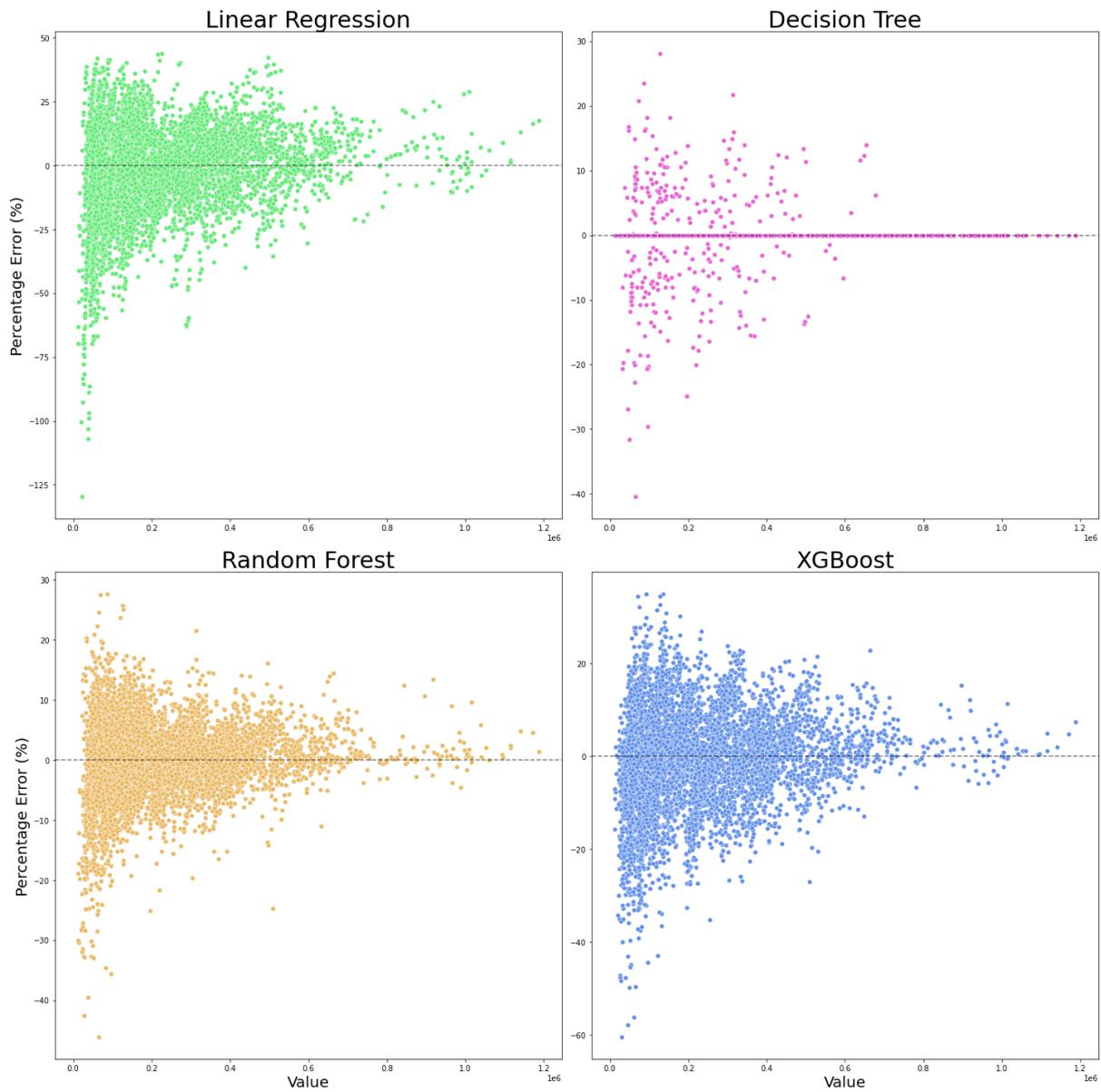
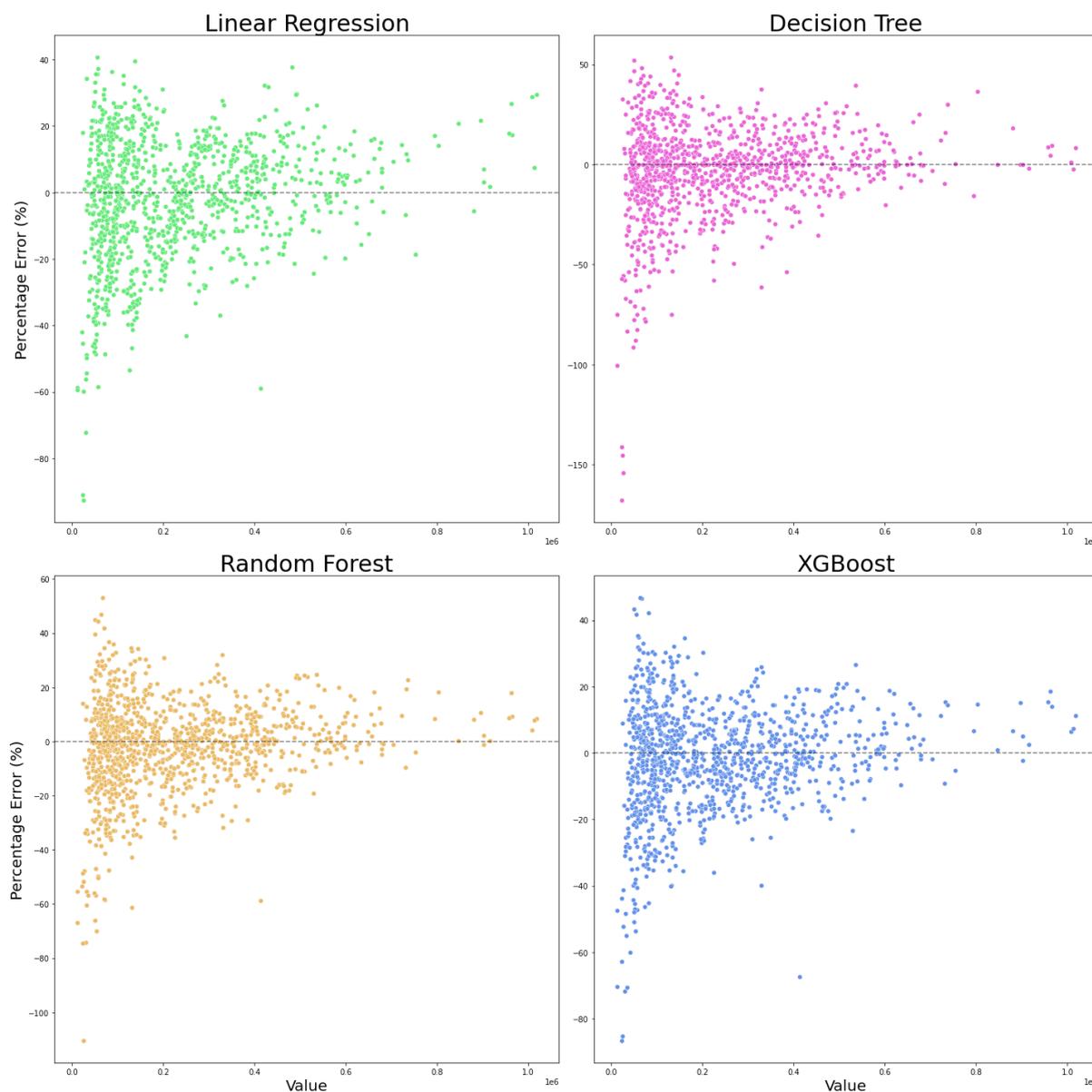


Figura 21. Error porcentual en evaluación.



Luego de analizar los resultados de los modelos en las métricas y las figuras provistas, se puede llegar a dos conclusiones:

1. El modelo de Regresión Lineal presenta *underfitting*, ilustrado en los residuos absolutos más altos tanto sobre los datos de entrenamiento como de evaluación de las Figuras 18 y 19, el modelo no logra aprender los patrones importantes necesarios para generalizar bien a nuevas instancias.
2. El modelo del Árbol de Decisión presenta *overfitting*, que se puede apreciar si comparamos las Figuras correspondientes a los datos de entrenamiento (donde el modelo presenta un error mínimo en la valoración de los dispositivos) con las

correspondientes a los datos de evaluación (donde el rendimiento del modelo cae drásticamente).

Respecto a los modelos restantes, como ambos presentan resultados similares en las predicciones individuales de evaluación visibles en las figuras, evaluamos estadísticamente la capacidad predictiva de los modelos restantes de XGBoost y Random Forest. Para esto se utiliza el *test* de Diebold-Mariano descrito a continuación.

### 5.1.1. Test de Diebold-Mariano para la evaluación de la dominancia de un modelo

El test de Diebold-Mariano (DM) es una herramienta estadística utilizada para comparar la precisión predictiva de dos modelos mediante la evaluación de la significatividad en las diferencias de los errores que tienen las predicciones. Originalmente se diseñó para comparar la precisión de las predicciones de forecasting en series temporales pero también puede aplicarse en el contexto de los modelos de regresión.

Para realizarlo es necesario obtener las predicciones de ambos modelos para el mismo conjunto de evaluación, calcular los errores de predicción mediante la diferencia con el valor real, aplicar la función de pérdida, obtener la diferencia entre los errores de predicción de los dos modelos y calcular el promedio y la varianza de estas diferencias. Luego, se computa el estadístico DM que sigue una distribución t de Student bajo la hipótesis nula de que los dos modelos tienen la misma precisión predictiva. Como para el RMSE no podemos aplicar la raíz sobre los puntos individuales, utilizamos el criterio del MSE que consiste en elevar los errores al cuadrado. El criterio del MAPE corresponde al definido en la Sección 3.5.3, donde se divide el error por el actual y se guarda el valor absoluto.

La significatividad estadística es evaluada al nivel del 5%, esto significa que si el p-valor es menor que 0.05 rechazamos la hipótesis nula sugiriendo que existe una diferencia estadísticamente significativa en la precisión predictiva de los modelos. Según el criterio, se detallan los resultados obtenidos de la comparación entre los modelos de Random Forest y XGBoost:

- MSE: Estadístico DM = 0.7242882806622949, p-valor = 0.4690399002377823.
- MAPE: Estadístico DM = 0.4778556437305111, p-valor = 0.632846117720656.

Para ambos casos, no hay suficiente evidencia para concluir que existe una diferencia significativa en la precisión predictiva de los modelos ya que no se puede rechazar la hipótesis

nula. Arbitrariamente se define al modelo de XGBoost como el modelo baseline a comparar con modelos optimizados ya que tiene, en promedio, un error porcentual y desvío en las predicciones menor al conseguido con el Random Forest.

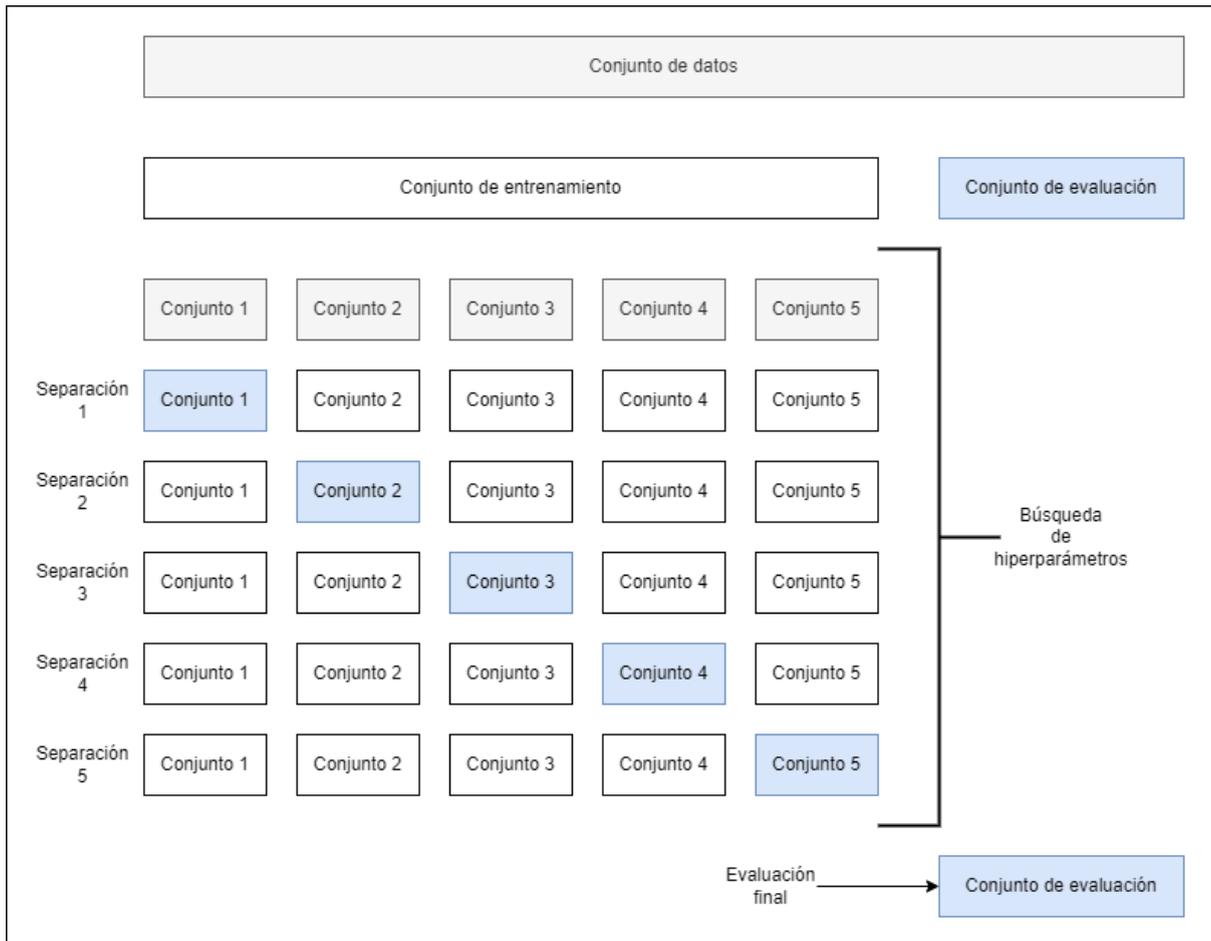
## 5.2. Evaluación de modelos ajustados

Los hiperparámetros son parámetros que no se aprenden automáticamente durante el entrenamiento del modelo. A diferencia de los parámetros internos (como los coeficientes en una regresión lineal), son establecidos antes de comenzar el proceso de entrenamiento, afectando directamente el rendimiento y la generalización del modelo. En esta sección comparamos nuestro modelo base con modelos cuyos hiperparámetros fueron ajustados mediante el método de búsqueda aleatoria con validación cruzada.

### 5.2.1. Búsqueda aleatoria de hiperparámetros con validación cruzada

Antes de entrenar nuestros modelos de predicción dividimos nuestros datos en conjuntos de entrenamiento y evaluación. Sin embargo, para encontrar la combinación óptima de hiperparámetros necesitamos de una muestra adicional denominada validación que nos permita evaluar el rendimiento del modelo hiperparametrizado con cada combinación. Como se ilustra en la Figura 22, en este proceso se separan conjuntos de entrenamiento y validación derivados del set de datos de entrenamiento original, que serán utilizados para entrenar y validar cada combinación elegida. En cada caso se evalúa el rendimiento promedio en los conjuntos de validación para elegir los hiperparámetros que maximizan el rendimiento del modelo.

Figura 22. Ilustración de la validación cruzada con una partición de 5.



Una de las formas más convencionales de encontrar los valores óptimos de los hiperparámetros es la búsqueda aleatoria, donde:

1. Se establece un espacio de búsqueda con los posibles valores para cada hiperparámetro.
2. Se seleccionan combinaciones de manera aleatoria dentro de cada espacio definido.
3. Se entrena el modelo y se evalúa el rendimiento con los datos de validación.
4. Se repite este proceso un número determinado de veces.
5. Se obtienen la mejor combinación de hiperparámetros encontrados.

La búsqueda aleatoria tiene ventajas sobre la búsqueda exhaustiva (donde probamos todas las combinaciones posibles) porque es más eficiente y se adapta bien a espacios extensos de hiperparámetros.

## 5.2.2. Hiperparámetros de los modelos

A continuación, se describen los hiperparámetros a ajustar de cada modelo, detallando entre corchetes el valor óptimo encontrado en la búsqueda aleatoria con validación cruzada.

Árbol de Decisión:

- “max\_depth”: Nos permite controlar la profundidad del árbol, limitarlo disminuye las probabilidades de sobreajustarse a los datos de entrenamiento [28].
- “min\_samples\_split”: Número mínimo de muestras necesarias para dividir un nodo interno, evitar los nodos con pocas muestras ayuda a prevenir el overfitting [4].
- “min\_samples\_leaf”: Número mínimo de muestras requeridas en un nodo terminal [5].
- “max\_features”: El número máximo de características a considerar para realizar la división en cada nodo [56].
- “criterion”: Especifica la función para medir la calidad de una división, entre las opciones disponibles aparecen el error cuadrático medio, el error absoluto medio, el error cuadrático medio de freidman y poisson ["squared\_error"].
- “splitter”: Especifica la estrategia utilizada en la división de cada nodo, donde la división aleatoria introduce mayor variabilidad que la denominada “mejor” ["random"].

Random Forest:

- “max\_features”: [42].
- “max\_depth”: [12].
- “min\_samples\_split”: [6].
- “min\_samples\_leaf”: [2].
- “criterion”: ["squared\_error"]
- “n\_estimators”: Determina el número de árboles en el bosque, aunque valores más altos ayudan a mejorar el rendimiento, también aumentan considerablemente el tiempo de entrenamiento [275].

XGBoost:

- “n\_estimators”: [425].
- “max\_depth”: [8].
- “min\_child\_weight”: [10].
- “learning\_rate”: Es la tasa de aprendizaje contribuida por cada árbol del modelo [0.02].
- “subsample”: La proporción de muestras a utilizar para ajustar cada árbol [0.775].
- “colsample\_bytree”: Proporción de características a utilizar para ajustar cada árbol [0.775].
- “gamma”: Representa la mínima reducción de pérdida necesaria para particionar un nodo [0.1].
- “reg\_alpha”: Parámetro de regularización L1 (Lasso) [0.07].
- “reg\_lambda”: Parámetro de regularización L2 (Ridge) [0.99].

### 5.2.3. Resultados de los modelos ajustados

La Tabla 14 detalla los resultados de los modelos ajustados y del modelo base a superar.

Tabla 14. Evaluación de modelos ajustados vs. *baseline*.

Model	Train R2	Test R2	Train RMSE	Test RMSE	Train MAPE	Test MAPE
XGBoost	98%	97%	21.780,68	32.211,75	7,54%	11,95%
Tunned Decision Tree	98%	96%	26.298,41	33.592,47	10,08%	12,17%
Tunned Random Forest	98%	97%	22.240,21	31.588,03	8,67%	11,62%
Tunned XGBoost	98%	97%	26.047,54	32.196,97	9,56%	11,71%

Analizamos los resultados de las métricas de los modelos ajustados:

- Vemos que el Árbol de Decisión ajustado ya no sufre del sobreajuste, explicando el 98% de la variabilidad del precio en los datos de entrenamiento y 96% en los de testeo, en promedio, su desvío de 26.298,41 pesos tiene un error porcentual absoluto del 10,08% en el entrenamiento, aumentando a \$33.592,47 y 2,09 puntos porcentuales en los datos de evaluación. Estos números no alcanzan a superar los obtenidos con el modelo base.
- El Random Forest ajustado consigue explicar 98% de la variabilidad con un desvío de \$22.240,21 y un error porcentual promedio de 8,67% en los datos de entrenamiento. Con los datos de evaluación, la explicación de la variabilidad cae en 1 punto porcentual, aumentando, en promedio, su desvío en precios hasta los \$31.588,03 y, en términos porcentuales, hasta 11,62%, superando al resto de los modelos evaluados.
- El XGBoost ajustado consigue durante el entrenamiento una explicación de la variabilidad del 98%, con un desvío en precio de \$26.047,54 y un error porcentual promedio del 9,56%. En la evaluación, la explicación de la variabilidad es del 97% con un desvío de 32.196,97 pesos y un error porcentual promedio del 11,71%, superando los resultados del modelo base, pero sin alcanzar los conseguidos con el modelo de bosque aleatorio.

Al igual que en la Sección 5.1.1, evaluamos con el *test* de Diebold-Mariano (DM) la precisión predictiva de los modelos de Random Forest y XGBoost ajustados, detallando según criterio:

- MSE: Estadístico DM = -1.0912032104776095, p-valor = 0.27541795429850857.
- MAPE: Estadístico DM = -0.8150140036954204, p-valor = 0.4152374094923106.

Para ambos casos, no hay suficiente evidencia para concluir que existe una diferencia significativa en la precisión predictiva de los modelos ya que no se puede rechazar la hipótesis nula, se visualizan en las Figuras 23, 24 y 25 los valores obtenidos por los modelos de ajustados en los datos de evaluación. Como se puede observar, ambos modelos tuvieron un comportamiento similar a la hora de predecir el precio de los celulares, donde (en general) los niveles de precio más bajo resultan en predicciones menos precisas que debieran considerarse a la hora de su implementación.

Es importante remarcar que en ambos casos, los modelos de predicción aparecen subvaluando los dispositivos con un nivel de precios más elevado, cuyo escenario es preferido si lo llevamos al caso donde el modelo es utilizado para la automatización de oferta de precios de compra a

clientes para las empresas. Por otro lado, cuando los celulares tienen un nivel de precio más bajo, el error en la predicción de ambos modelos llega a niveles más altos de sobrevaloración.

Figura 23. Valor actual vs. predicho con línea de predicción perfecta en evaluación.

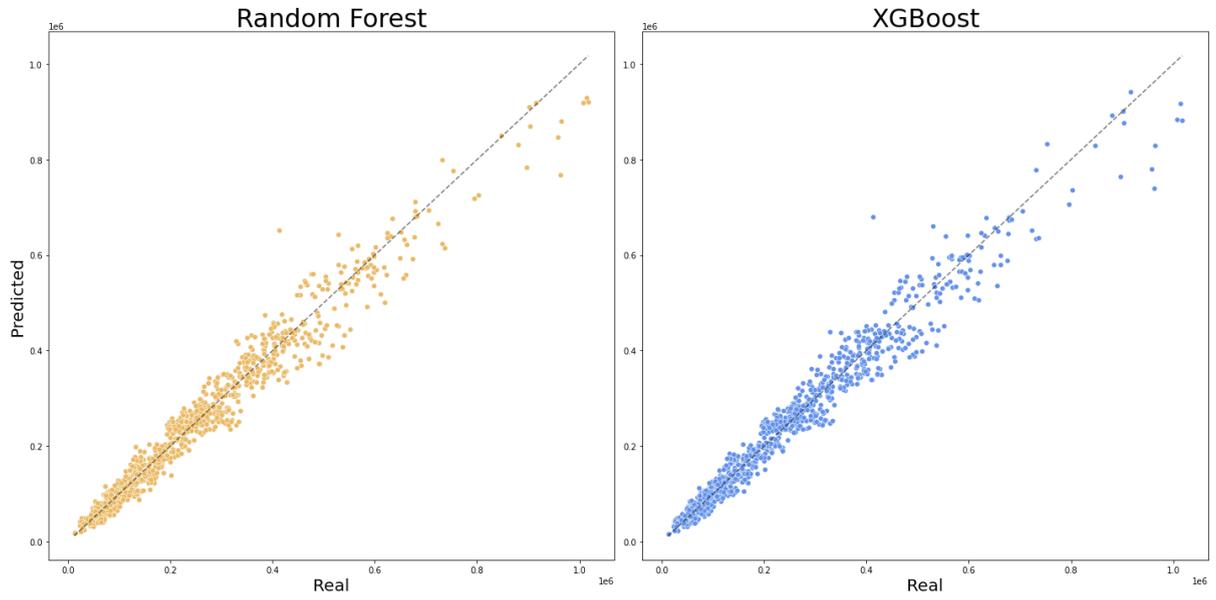


Figura 24. Residuos en evaluación.

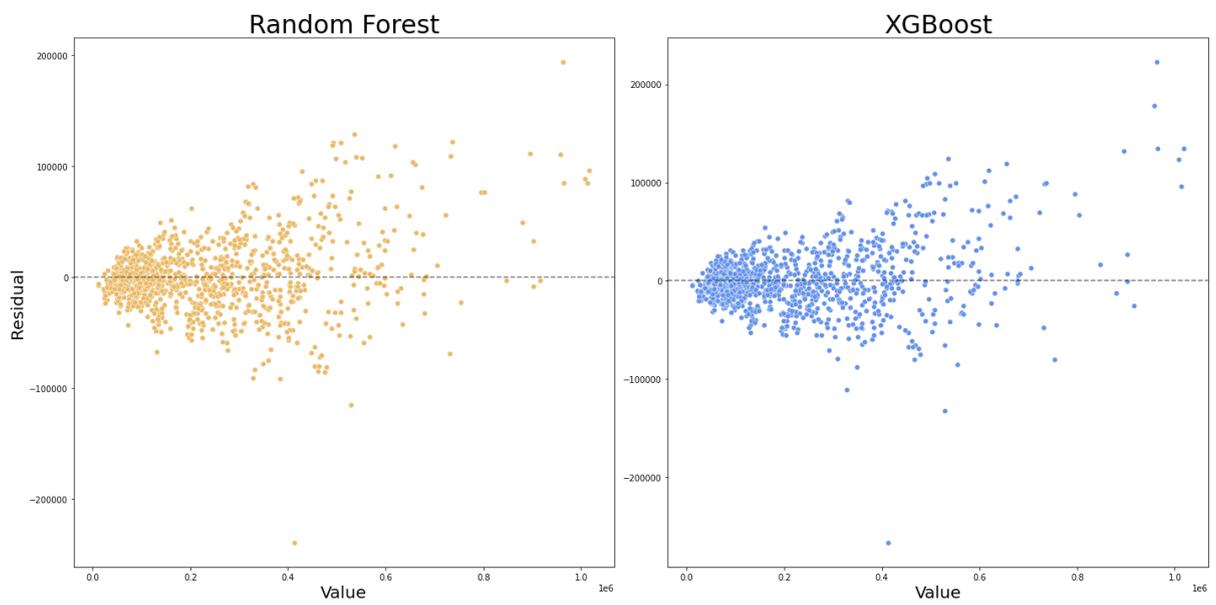
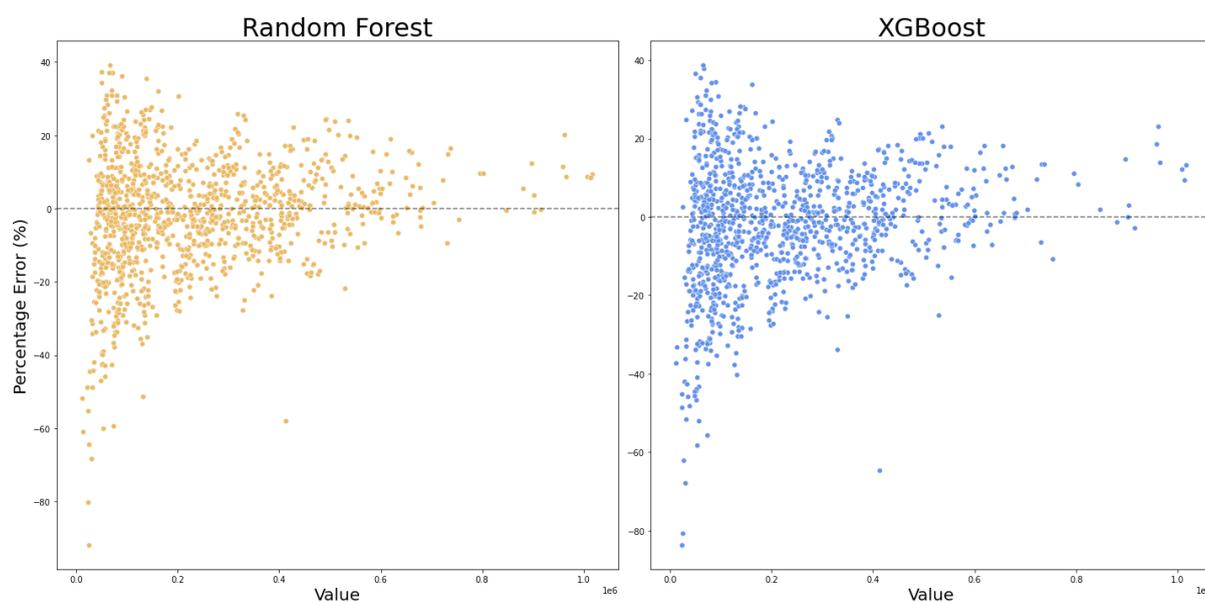


Figura 25. Error porcentual en evaluación.



Utilizando las predicciones del modelo de Random Forest, en la Tabla 1 del anexo se detallan las estadísticas descriptivas de los residuos en nuestros datos de evaluación para cada combinación de marca, modelo, capacidad y condición. Subsecuentemente, la Tabla 2 del anexo detalla la potencial ganancia porcentual que podría obtener una empresa con la posibilidad de esperar hasta 30 días para vender el dispositivo comprado al cliente, donde asumimos que:

- El tiempo de venta del cliente es igual a 1, mientras que el de la empresa es igual a 30 (“time\_to\_sell”).
- La calidad de las imágenes del cliente es mala, mientras que las de la empresa buena (“tag\_poor\_quality\_picture” / “tag\_good\_quality\_picture” / “tag\_good\_quality\_thumbnail”).
- El cliente no ofrece garantías en el producto ni en su envío, mientras que la empresa sí (“tag\_extended\_warranty\_eligible” / “tag\_shipping\_guaranteed”).

En los atributos restantes se mantiene el mismo valor para las predicciones de los grupos. Evaluar las predicciones a este nivel nos permite determinar cuales son las capacidades reales que tiene el modelo ya que fue entrenado sobre datos que no contemplan la diferentes condiciones en las que se encuentran los dispositivos (un usado podría tener simplemente una sola semana de uso, mientras que el otro 2 años), y presentan la susceptibilidad de la información provista por los usuarios dentro de la plataforma (que catalogan nuevos como usados, o usados como reacondicionados).

Si consideramos todos los dispositivos, el promedio de ganancia porcentual encontrado es del 12,52%, con un desvío estándar de 8,28 puntos porcentuales, un mínimo de -2,16% (observando en detalle se encuentran 8 celulares que potencialmente pierden valor en el mercado con el paso del tiempo, aunque pertenecen a los modelos que menos registros presentan) y un máximo de 43,99% correspondiente al iPhone 7 de Apple. Recordando que un 0 indica una condición de usado, 1 de reacondicionado y 2 nuevo, en las tablas 15 y 16 se detallan estas estadísticas desglosadas por marca y condición respectivamente.

Tabla 15. Estadísticas descriptivas del potencial porcentaje de ganancia por marca.

Marca	count	mean	std	min	max
Apple	73	13,02%	8,25	-0,56%	43,99%
Motorola	16	13,53%	9,75	-2,16%	30,36%
Samsung	43	11,79%	8,19	-1,85%	26,22%
Xiaomi	7	9,58%	5,94	-0,20%	17,39%

Tabla 16. Estadísticas descriptivas del potencial porcentaje de ganancia por condición.

item_condition	count	mean	std	min	max
0	89	13,87%	8,97	-1,85%	43,99%
1	13	13,12%	7,71	0,29%	24,89%
2	37	9,07%	5,44	-2,16%	23,32%

Considerando que la empresa puede reacondicionar el dispositivo usado antes de venderlo, detallamos en la Tabla 3 del anexo la potencial ganancia bruta porcentual (que no considera los costos asociados al reacondicionamiento) predicha por el modelo, donde encontramos que el promedio sube hasta los 43,09% con un desvío estándar de 25,14 puntos porcentuales, mínimo de -6,56% correspondiente al Samsung Galaxy S22 Ultra 5G (que aparece siendo subvaluado por el modelo, con un residuo promedio de \$51.549,85, desvío de \$40.056,91 pesos y mínimos y máximos de -95.072,03 y -1.965,98 respectivamente), y un máximo de 87,12% correspondiente, de nuevo, con el iPhone 7. En la Tabla 17 se muestran estas estadísticas para las marcas Apple y Samsung.

Tabla 17. Estadísticas descriptivas del potencial porcentaje de ganancia bruta por reacondicionamiento y espera desglosado por marca.

Marca	count	mean	std	min	max
Apple	6	61,29%	16,81	43,63%	87,12%
Samsung	6	24,89%	17,67	-6,56%	40,93%

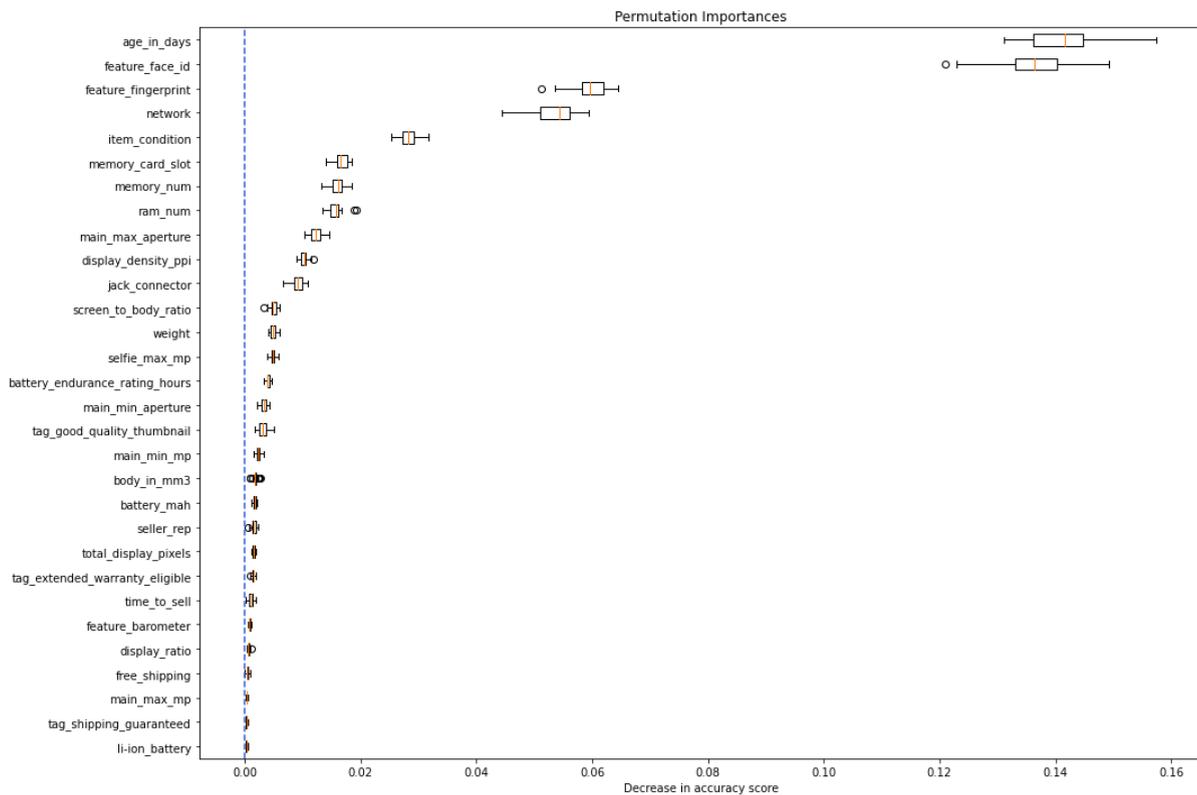
Se destaca la valoración que da el mercado al reacondicionamiento de los celulares Apple, que presentan una diferencia positiva de 36,5 puntos porcentuales en la ganancia promedio obtenida por el reacondicionamiento. En la sección siguiente analizamos la influencia de los distintos atributos en los resultados obtenidos mediante el cálculo de la importancia de atributos con permutación.

#### 5.2.4. Importancia de los atributos del modelo

Explorando la importancia de los atributos podemos identificar cuales son las características que más afectan nuestras predicciones para poder comprender mejor las capacidades y limitaciones. Se decidió utilizar el método permutativo, que es una técnica que mide la contribución de cada atributo del modelo entrenado reordenando aleatoriamente los valores de cada *feature* individual y observando la degradación resultante en su capacidad predictiva.

Entre las variables más importantes en términos de crecimiento de la capacidad predictiva del modelo detalladas en la Figura 26 destacamos las que preponderan en la marca Apple, “feature\_face\_id”, “feature\_fingerprint”, “memory\_card\_slot” y “jack\_connector”, ya que visibilizan la diferenciación en el precio que tienen los celulares cuya marca es valorada por los usuarios (y la diferenciación que tienen las innovaciones tecnológicas implementadas).

Figura 26. Primeros 30 atributos más importantes del modelo según el método permutativo.



A continuación, se entrena un modelo predictivo con el dispositivo de mayor frecuencia en nuestros datos de entrenamiento (Apple iPhone 11) para comparar la capacidad predictiva que tiene nuestro modelo general contra uno individual y entender qué variables influyen en las predicciones en ese nivel.

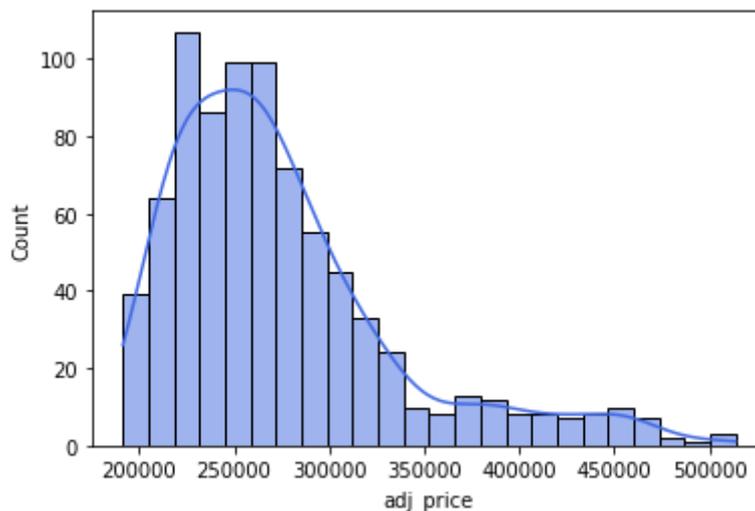
### 5.2.5. Comparación contra modelo individual de dispositivo

En esta sección filtramos en nuestros datos de entrenamiento y evaluación el dispositivo con mayor frecuencia de entrenamiento, el iPhone 11 de Apple (con 820 registros de entrenamiento), para comparar los resultados obtenidos por un modelo capaz de predecir distintos dispositivos en simultáneo, contra las de un modelo entrenado para predecir un solo celular del mercado.

Investigar la diferencia en las capacidades predictivas de ambos modelos nos ayuda a entender mejor la solución, que pudiera requerir del desarrollo de distintos modelos de predicción para cada dispositivo debido a sus diferencias naturales. Al utilizar un solo dispositivo para entrenar un modelo de *machine learning* se puede capturar mejor las características que influyen en su precio, obteniendo una mejor comprensión de aquellas que son más importantes y resultando en predicciones más precisas para el dispositivo en particular.

Se eliminan algunas columnas ya que no se encuentra diferenciación en los registros filtrados para el mismo dispositivo: "tag\_3x\_campaign", "tag\_ahora-3", "tag\_ahora-6", "tag\_ahora-paid-by-buyer", "tag\_cart\_eligible", "tag\_catalog\_boost", "tag\_catalog\_only\_restricted", "tag\_deal\_of\_the\_day", "tag\_meli\_choice\_candidate", "tag\_mshops\_3x\_campaign", "ram\_num", "network", "weight", "nfc", "radio", "status\_available", "body\_in\_mm3", "screen\_to\_body\_ratio", "total\_display\_pixels", "display\_ratio", "display\_density\_ppi", "cpu\_cores", "memory\_card\_slot", "main\_cameras", "main\_min\_mp", "main\_max\_mp", "main\_min\_aperture", "main\_max\_aperture", "selfie\_cameras", "selfie\_max\_mp", "jack\_connector", "feature\_fingerprint", "feature\_gyro", "feature\_proximity", "feature\_compass", "feature\_barometer", "feature\_face\_id", "feature\_iris\_scanner", "feature\_heart\_rate", "feature\_spo2", "li-ion\_battery", "battery\_mah", "battery\_endurance\_rating\_hours", "brand\_Apple", "brand\_Motorola", "brand\_Samsung". Visualizamos la distribución de la variable objetivo en la Figura 27, decidimos no aplicar ninguna transformación ya que no presenta una falta de simetría problemática como en los datos que incorporan a todos los dispositivos.

Figura 27. Histograma de la variable objetivo en los datos de entrenamiento filtrados.



En la Tabla 18 continuamos el desarrollo del modelo evaluando los resultados en el entrenamiento y evaluación para los distintos modelos base, donde definimos a la regresión lineal como *baseline* a superar por los modelos ajustados.

Tabla 18. Evaluación de los Modelos de Predicción sin Hiperparametrización.

Model	Train R2	Test R2	Train RMSE	Test RMSE	Train MAPE	Test MAPE
Linear Regression	73%	75%	31.705,37	34.748,27	9,79%	10,56%
Decision Tree	99%	63%	5.864,97	42.388,70	0,47%	12,44%
Random Forest	95%	74%	13.814,16	35.890,03	4,07%	10,40%
XGBoost	97%	74%	9.900,28	36.072,73	2,53%	10,85%

Analizamos los resultados de las métricas de cada modelo:

- La Regresión Lineal explica el 73% de la variabilidad del precio en el entrenamiento, en la evaluación encontramos que la explicabilidad sube en 2 puntos porcentuales y, en promedio, durante el entrenamiento presenta un desvío de 31.705,37 pesos que representa un error porcentual del 9,79%, mientras que en los datos de evaluación el desvío se incrementa hasta los \$34.748,27, aumentando también el error porcentual hasta 10,56%. A diferencia del modelo general, el modelo no presenta indicios de subajuste en los datos filtrados.
- En el Árbol de Decisión notamos nuevamente indicios de sobreajuste ya que, durante el entrenamiento, el modelo logra explicar casi toda la variabilidad del precio teniendo un desvío promedio de apenas \$5.864,97 que representan un error porcentual promedio de 0,47%, pero estos resultados no son replicados en la evaluación, donde la variabilidad explicada cae 36 puntos porcentuales y, en promedio, el desvío es de 42.388,70 pesos que representan un error porcentual de 12,44%.
- El modelo de Bosque Aleatorio pareciera estar sobreajustándose (en menor medida que el árbol de decisión) a los datos ya que logra explicar un 95% de la variabilidad en entrenamiento con, en promedio, 13.814,16 pesos de desvío y un error porcentual de 4,07%. En la evaluación, la explicación de la variabilidad cae hasta el 74% y, en promedio, el desvío en precios sube hasta los \$35.890,03, con un error porcentual de 10,40%.

- Al igual que con el modelo de bosque aleatorio, el XGBoost presenta indicios de sobreajuste ya que consigue explicar el 99% de la variabilidad en el precio, con un desvío promedio de estos en 9.900,28 pesos y un error porcentual de 2,53%. En la evaluación, la explicación de la variabilidad cae en 23 puntos porcentuales mientras que, en promedio, el desvío sube hasta los 36.072,73 pesos y el error porcentual alcanza el 10.85%.

Luego de ajustar los modelos de Árbol de Decisión, Bosque Aleatorio y XGBoost mediante la búsqueda aleatoria de hiperparametros con validación cruzada, evaluamos los resultados contra el modelo base a superar en la Tabla 19.

Tabla 19. Evaluación de Modelos Ajustados vs. *Baseline*.

Model	Train R2	Test R2	Train RMSE	Test RMSE	Train MAPE	Test MAPE
Linear Regression	73%	75%	31.705,37	34.748,27	9,79%	10,56%
Tunned Decision Tree	72%	72%	31.968,81	36.798,03	9,49%	10,97%
Tunned Random Forest	83%	74%	25.282,94	35.656,24	7,35%	10,30%
Tunned XGBoost	81%	76%	26.152,50	34.172,49	8,01%	10,22%

Analizamos los resultados de las métricas de los modelos ajustados que ya no muestran indicios de sobreajuste:

- Vemos que el Árbol de Decisión explica el 72% de la variabilidad del precio en los datos de entrenamiento y 72% en los de testeo, en promedio, su desvío de 31.968,81 pesos tiene un error porcentual absoluto del 9,49% en el entrenamiento, aumentando a \$36.798,03 y 1,48 puntos porcentuales en los datos de evaluación. Estos números no alcanzan a superar los obtenidos con el modelo base.
- El Random Forest ajustado consigue explicar 83% de la variabilidad con un desvío de \$25.282,94 y un error porcentual promedio de 7,35% en los datos de entrenamiento. Con los datos de evaluación, la explicación de la variabilidad cae en 9 puntos

porcentuales, aumentando, en promedio, su desvío en precios hasta los \$35.656,24 y, en términos porcentuales, hasta 10,30%.

- El XGBoost ajustado consigue durante el entrenamiento una explicación de la variabilidad del 81%, con un desvío en precio de \$26.152,50 y un error porcentual promedio del 8,01%. En la evaluación, la explicación de la variabilidad es del 76% con un desvío de 34.172,49 pesos y un error porcentual promedio del 10,22%, superando los resultados del modelo base y del bosque aleatorio.

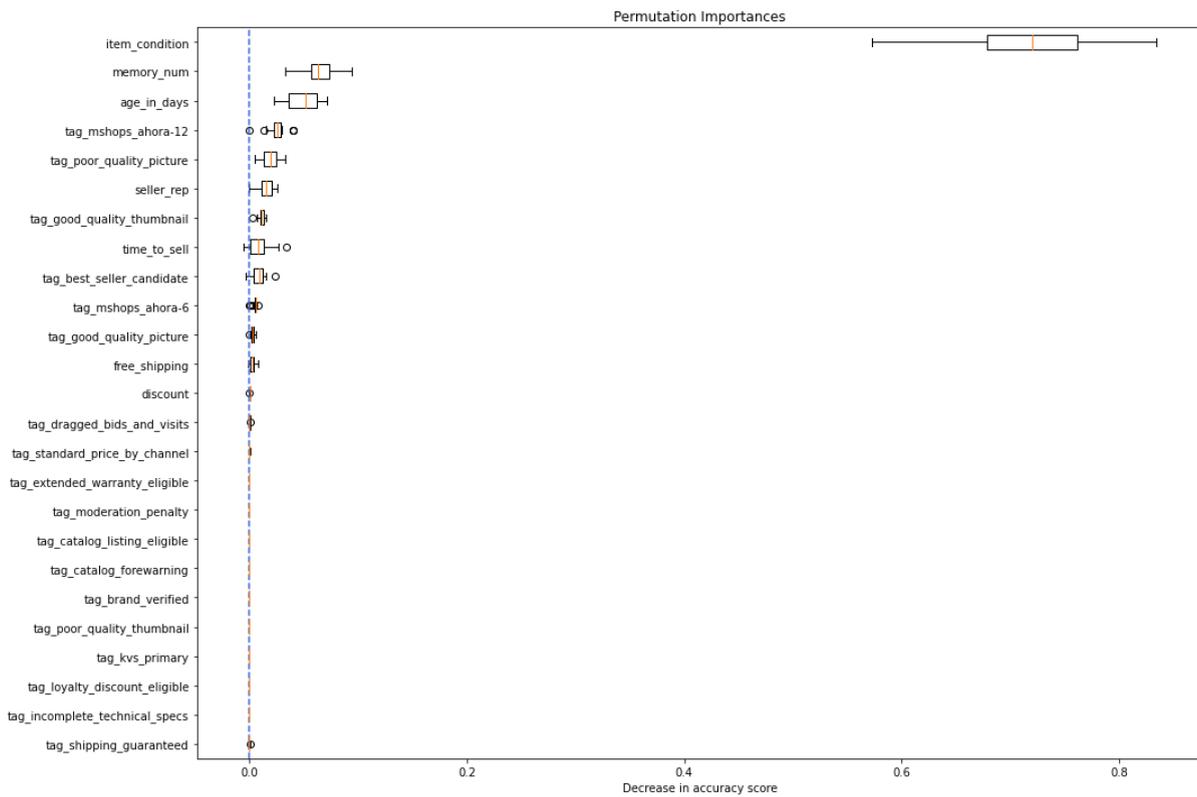
Como el modelo de Random Forest aparece muy cerca del de XGBoost en el error porcentual promedio, evaluamos con el *test* de Diebold-Mariano (DM) su precisión predictiva, detallando según criterio:

- MSE: Estadístico DM = 2.387968115964217, p-valor = 0.018266821670258487.
- MAPE: Estadístico DM = 0.4534417712370366, p-valor = 0.6509277409911982.

Vemos que para el error porcentual absoluto promedio no hay suficiente evidencia para concluir que existe una diferencia significativa en la precisión predictiva de los modelos ya que no se puede rechazar la hipótesis nula. Pero este no es el caso en términos del error cuadrático promedio, donde rechazamos la hipótesis nula asegurando que existe una diferencia significativa en la precisión predictiva del XGBoost.

En la Figura 28 se visualiza la importancia de los atributos por permutación del modelo ajustado de XGBoost, donde las variables que más afectan los resultados del modelo son la condición, capacidad de memoria, edad al momento de compra, posibilidad de cuotas, calidad de imagen, reputación del vendedor y tiempo de espera hasta la venta.

Figura 28. Importancia de atributos del modelo individual según el método permutativo.



A continuación, utilizamos los valores detallados para el iPhone 11 de Apple en las Tablas 1, 2 y 3 del Anexo y los comparamos con los obtenidos por el modelo individual desarrollado. La Tabla 20 detalla las estadísticas descriptivas de los residuos de cada modelo, donde se observa que los residuos de las predicciones de un modelo entrenado solamente para un celular presentan un menor desvío, demostrando ser más consistente a la hora de predecir el precio. Sin embargo, si consideramos que la implementación del modelo puede ser supervisada por expertos en el área que definan los precios máximos y mínimos para cada dispositivo dadas capacidades y condiciones, creemos que la diferenciación en la capacidad de los mismos no es suficiente como para elegirlo por sobre el general.

Tabla 20. Estadísticas descriptivas de los residuos del iPhone 11 por modelo, capacidad y condición.

memory _num	item_co ndition	count	residual_general_model				residual_iphone_11_model			
			mean	std	min	max	mean	std	min	max
64	0	57	8.064,07	26.118,39	-47.572,72	51.397,02	8.161,40	26.946,60	-50.696,65	52.737,88
64	1	1	20.564,71		20.564,71	20.564,71	21.758,28		21.758,28	21.758,28
64	2	11	2.687,54	16.918,77	-20.630,43	21.531,47	359,08	12.972,00	-17.793,56	16.112,36
128	0	53	-3.306,64	37.149,00	-84.007,25	56.659,66	-1.414,08	37.555,73	-89.073,38	56.521,02
128	1	6	-55.537,94	67.563,71	-120.983,13	41.602,26	-39.227,04	67.454,86	-111.496,58	53.405,52
128	2	6	-9.297,03	25.226,24	-53.889,31	11.089,86	-7.486,24	21.608,85	-41.505,10	11.874,12
256	0	8	-3.243,16	39.629,73	-66.458,82	42.974,51	2.621,32	39.726,02	-64.571,78	54.251,40

En la Tabla 21 se detalla la potencial ganancia porcentual que obtendría una empresa con la posibilidad de esperar hasta 30 días para vender el dispositivo comprado al cliente desglosado por su condición y capacidad. Se agrega una condición tipo “3” que ilustra la ganancia porcentual bruta de la compra del dispositivo usado con el objetivo de reacondicionarlo antes de venderlo. En la misma se observa que el modelo individual, salvo capacidades altas de memoria, es más pesimista a la hora de proyectar la potencial ganancia porcentual de los celulares usados, pero muestra optimismo en los casos donde este es reacondicionado.

Tabla 21. Potencial ganancia porcentual del iPhone 11 por modelo, capacidad y condición.

memory _num	item_co ndition	general_model			iphone_11_model		
		pred_1	pred_30	%profit	pred_1	pred_30	%profit
64	0	225.806,82	281.561,81	24,69	223.760,31	259.716,00	16,07
128	0	245.590,24	287.753,99	17,17	243.505,77	270.249,88	10,98
256	0	237.949,39	297.015,70	24,82	226.195,39	278.331,47	23,05
64	1	342.836,23	386.602,45	12,77	314.141,34	389.956,38	24,13
128	1	342.068,07	374.170,23	9,38	316.864,91	376.505,81	18,82
64	2	340.327,97	376.411,70	10,60	328.752,44	385.564,97	17,28
128	2	350.424,49	374.635,14	6,91	343.699,41	383.431,94	11,56
64	3	225.806,82	386.602,45	71,21	223.760,31	389.956,38	74,27
128	3	245.590,24	374.170,23	52,36	243.505,77	376.505,81	54,62

## 6. Consideraciones generales

En esta sección se examinan los supuestos subyacentes en los datos que afectan las capacidades predictivas del modelo, desarrollando las extensiones para corregirlos y proponiendo una implementación de la solución para una empresa dedicada a la compraventa de celulares que busca aumentar la fidelidad del cliente mediante la solución de un problema común que estos tienen a la hora de actualizar su celular (¿qué hacer con el viejo?).

### 6.1. Supuestos subyacentes en los datos

En la Sección 4.2, a la hora de diseñar la estructura de la información que se utiliza para entrenar nuestro modelo predictivo, se asume que aquellos registros que desaparecen de la plataforma representan una venta al último precio en el que estaban publicados. Este precio asignado por los vendedores considera los costos de publicación e impuestos dentro de la plataforma, por lo que es razonable que un vendedor que opte por vender el dispositivo fuera de Mercado Libre muestre disposición a disminuir el valor de la transacción entre las partes.

La condición de los dispositivos es asignada por los vendedores, por lo que es posible que existan inconsistencias donde, por ejemplo, un usado sea catalogado como nuevo o reacondicionado a un precio que no sea identificado como outlier (ver Sección 4.4.3), afectando la calidad de nuestras predicciones. Además de esta problemática, las categorías “Reacondicionado” y “Usado” no definen de manera suficiente la naturaleza del producto, ya que, por ejemplo, un celular usado es usado con sólo una semana de uso mientras que otros pueden tener varios años e incluso ya no funcionar debido al desgaste o algún daño.

### 6.2. Extensiones

Entendiendo que la mayor falencia del modelo es la condición real en la que se encuentran los dispositivos, se recomienda apoyar la solución con el desarrollo de modelos adicionales que nos ayuden a identificar y catalogar las distintas variaciones que tenemos de las condiciones. Uno capaz de identificar el celular y su estado visible a través del entrenamiento con imágenes de los distintos celulares en diferentes condiciones, y otro entrenado para predecir el estado del celular en base a descripciones provistas por el usuario. Creemos que uno no es excluyente del otro ya que existe la posibilidad de que, por ejemplo, la batería del celular se descargue más rápido de lo esperado por el mismo desgaste de su uso, y esto solo pueda ser identificado en la descripción provista por el usuario. En base al resultado obtenido por ambos modelos se

debería catalogar al dispositivo en una categoría definida por el negocio que luego se utiliza como atributo en el modelo de predicción de precio de mercado.

Para el entrenamiento de estos modelos podrían utilizarse las descripciones, comentarios e imágenes correspondientes a las publicaciones dentro de Mercado Libre, con la inclusión de la condición real como atributo, el modelo tendría la capacidad de diferenciar las mismas a un nivel más granular, pudiendo asignar con mayor certidumbre un precio de oferta. Si la determinación del estado del dispositivo fuera realizada sobre otra base de datos (descripciones generadas con modelos de lenguaje, imágenes de celulares en distintas condiciones identificadas manualmente por la empresa, etc.), la empresa debería apalancarse más de un expertise técnico, revisando las ofertas realizadas para cada modelo, identificando el estado al que corresponde su predicción, y asignando un descuento/incremento a su valor siguiendo un criterio representativo de los distintos estados.

Por otro lado, en este trabajo se asigna arbitrariamente un tiempo de reventa de 30 días que en realidad aparece como variable independiente en el modelo. Futuros desarrollos podrían incluir diferentes escenarios de días acordes a las necesidades específicas de una empresa o a la optimización a través de la maximización del beneficio presente en base al precio predicho para cada día y una tasa de descuento que permita comparar los diferentes valores.

### 6.3. Implementación propuesta

Asumiendo que la empresa quiere empezar a implementar la automatización de ofertas de compra de celulares a clientes con el modelo propuesto, se recomienda que ésta sea focalizada en el cliente, permitiendo feedback acerca de la oferta o sobre celulares no disponibles para su evaluación manual por el equipo de soporte. En el proceso, como se ilustra en la Figura 29, es importante requerir al usuario que ingrese la información correspondiente al dispositivo para recopilar nuevos registros que permitan nutrir al modelo con más información acerca de los celulares (pudiendo, por ejemplo, incluir nuevos modelos no contemplados en el modelo o que se encuentran actualmente filtrados por el equipo).

En su implementación, si consideramos que la oferta se realiza con crédito interno (dinero que el usuario solo puede utilizar dentro de la plataforma para comprar un celular nuevo, fundas, cargadores, etc.) la empresa podría realizar un aumento del valor asignado por el modelo, recompensando a los clientes por su lealtad continua hacia la empresa y motivándolos a elegirla cada vez que quieran actualizar su celular.

Figura 29. Ilustración de la implementación propuesta.



## 7. Conclusiones

En el desarrollo de este trabajo se detalló la implementación punta a punta de un modelo de *machine learning* diseñado para implementar una solución a una falencia en el mercado de celulares en Argentina.

Primero, se identificó la oportunidad que tienen los celulares usados en el país, donde la mayoría de los agentes corresponden a individuos que asumen riesgos que una empresa puede capitalizar construyendo una marca alrededor del segmento, y agregando aún más valor mediante el reacondicionamiento. En segundo lugar, se recolectó la información disponible vía API en el *marketplace* más importante del país, Mercado Libre, para diseñar una fuente de datos que refleje los precios y tiempos de transacción dentro del mercado. Se resalta la importancia que necesita este proceso ya que es la fuente de información con la cual se alimenta al modelo que luego toma decisiones en las unidades del negocio, y generalmente, como sucede en este caso, presenta desafíos. Al encontrar que la información de los atributos no estaba estandarizada dentro de la plataforma, se identificaron manualmente las distintas combinaciones de dispositivos encontrados (utilizando la información provista por el vendedor en los títulos y atributos de la publicación) y se realizó una segunda extracción en una fuente de datos donde los atributos de los celulares están estandarizados con *web scraping*, GSMarena.

Luego de identificar cada modelo de celular e integrar las dos fuentes de datos, se ajustó para cada día el precio de los dispositivos con el fin de eliminar el efecto inflacionario del país a través de la interpolación lineal del Índice de Precios al Consumidor, calculado mensualmente por el Instituto Nacional de Estadística y Censos de la República Argentina. Se eliminaron registros atípicos cuyos precios exceden los comprendidos en el mercado y se calcularon los rangos intercuartiles para identificar y eliminar los *outliers* en cada combinación de celular, capacidad y condición.

Con la información resultante se entrenó un modelo de regresión capaz de predecir los precios de los distintos dispositivos dados atributos intrínsecos del vendedor como la reputación desarrollada y tiempos máximos de venta en simultáneo, este fue comparado con un modelo entrenado solo con un dispositivo (que tiene la mayor cantidad de observaciones registradas) para evaluar las capacidades y obtener información respecto de las predicciones en cada caso. En el proceso encontramos que, con un tiempo de espera máximo de 30 días hasta la venta, existe una oportunidad para las empresas de Argentina de sacar ventaja de la actual ineficiencia presente en el mercado de celulares, la cual se extiende incluso a la posibilidad de optimizar la

rentabilidad esperada con tiempos de espera dinámicos para cada dispositivo que considere, entre otras cosas, los costos de depreciación e inventario.

Se resalta la importancia de evaluar la calidad de los resultados para cada modelo ya que, aunque se intentó extraerlo de los títulos (pero la información no era suficientemente representativa), las condiciones de “Usado” y “Reacondicionado” no reflejan al detalle la naturaleza de los productos publicados. Por esto, se sugiere que en futuras líneas de investigación se incluyan descripciones, comentarios e imágenes correspondientes a las publicaciones. Si se quisiera implementar la solución, se recomienda de un exhaustivo análisis experto de las predicciones para identificar la naturaleza de cada modelo encontrado en los datos, pudiendo así determinar las modificaciones de los precios para los estados no identificados.

# Referencias

1. [Abhishek, Kumar, Awasthi., Jinhui, Li., Lenny, Koh., Oladele, A., Ogunseitan. \(2019\). Circular economy and electronic waste. Nature Electronics, 2\(3\):86-89.](#)
2. [Chandrashekhara, K.T., Thungamani, M., Gireesh Babu, C.N., Manjunath, T.N. \(2019\). Smartphone Price Prediction in Retail Industry Using Machine Learning Techniques. In: Sridhar, V., Padma, M., Rao, K. \(eds\) Emerging Research in Electronics, Computer Science and Technology. Lecture Notes in Electrical Engineering, vol 545. Springer, Singapore. https://doi.org/10.1007/978-981-13-5802-9\\_34](#)
3. [H. Liu, J. Huang, H. Han and H. Yang, "An Improved Intelligent Pricing Model for Recycled Mobile Phones." 2020 Chinese Automation Congress \(CAC\), Shanghai, China, 2020, pp. 3724-3731](#)
4. [Joshua Bloch. 2006. How to design a good API and why it matters. In Companion to the 21st ACM SIGPLAN symposium on Object-oriented programming systems, languages, and applications \(OOPSLA '06\). Association for Computing Machinery, New York, NY, USA, 506-507.](#)
5. [Liang, Qipeng. \(2024\). Mobile phone price prediction: A comparative study among four models. Applied and Computational Engineering. 48. 212-218. 10.54254/2755-2721/48/20241516.](#)
6. [Liangchuan, Zhou., Surendra, M., Gupta. \(2019\). Marketing research and life cycle pricing strategies for new and remanufactured products.](#)
7. [M. Çetin and Y. Koç, "Mobile Phone Price Class Prediction Using Different Classification Algorithms with Feature Selection and Parameter Optimization," 2021 5th International Symposium on Multidisciplinary Studies and Innovative Technologies \(ISMSIT\), Ankara, Turkey, 2021, pp. 483-487](#)
8. [Maesya, Aries et. al., \(2023\). Mobile Phone Price Prediction Based on Supervised Learning Algorithms. International Journal of Applied Engineering and Technology 5\(1\), pp 41-44.](#)
9. [Muhammad Asim, Zafar Khan . Mobile Price Class prediction using Machine Learning Techniques. International Journal of Computer Applications. 179, 29 \( Mar 2018\), 6-11.](#)

10. [Ocampo, E. \(2017\). Fighting inflation in argentina: a brief history of ten stabilization plans.](#)
11. [Ravi Subramanian, Ramanath Subramanyam, \(2012\). Key Factors in the Market for Remanufactured Products](#)
12. [Tamar, Makov, Tomer, Fishman., Marian, Chertow., Vered, Blass. \(2018\). What Affects the Secondhand Value of Smartphones: Evidence from eBay](#)
13. [Yacan, Wang., Benjamin, T. Hazen. \(2016\). Consumer product knowledge and intention to purchase remanufactured products](#)
14. [Zehtab-Salmasi, A., Feizi-Derakhshi, AR., Nikzad-Khasmakhi, N. et al. Multimodal Price Prediction. Ann. Data. Sci. 10, 619–635 \(2023\).](#)
15. [Zhou, L., & Gupta, S. M. \(2020\). Value depreciation factors for new and remanufactured high-technology products: a case study on iPhones and iPads.](#)

# Anexo

Tabla 1. Estadísticas descriptivas de los residuos por dispositivo, capacidad y condición.

Marca	Modelo	ram_num	memory_num	item_condition	residual				
					count	mean	std	min	max
Apple	iPhone 11	4,00	64,00	0,00	57,00	8.064,07	26.118,39	-47.572,72	51.397,02
Apple	iPhone 11	4,00	64,00	1,00	1,00	20.564,71		20.564,71	20.564,71
Apple	iPhone 11	4,00	64,00	2,00	11,00	2.687,54	16.918,77	-20.630,43	21.531,47
Apple	iPhone 11	4,00	128,00	0,00	53,00	-3.306,64	37.149,00	-84.007,25	56.659,66
Apple	iPhone 11	4,00	128,00	1,00	6,00	-55.537,94	67.563,71	-120.983,13	41.602,26
Apple	iPhone 11	4,00	128,00	2,00	6,00	-9.297,03	25.226,24	-53.889,31	11.089,86
Apple	iPhone 11	4,00	256,00	0,00	8,00	-3.243,16	39.629,73	-66.458,82	42.974,51
Apple	iPhone 11 Pro	4,00	64,00	0,00	8,00	-6.904,77	30.922,19	-62.354,76	25.501,02
Apple	iPhone 11 Pro	4,00	64,00	1,00	6,00	-15.388,61	68.450,32	-128.972,19	50.399,03
Apple	iPhone 11 Pro	4,00	256,00	0,00	15,00	5.806,48	32.862,26	-54.057,70	58.736,95
Apple	iPhone 11 Pro Max	4,00	64,00	0,00	12,00	7.893,54	31.990,79	-37.757,59	65.701,15
Apple	iPhone 11 Pro Max	4,00	256,00	0,00	9,00	-626,90	44.671,46	-71.826,82	40.882,23
Apple	iPhone 12	4,00	64,00	0,00	13,00	3.461,56	21.925,04	-35.014,28	34.286,87
Apple	iPhone 12	4,00	64,00	1,00	6,00	44.348,10	36.644,82	-20.828,62	91.144,97
Apple	iPhone 12	4,00	64,00	2,00	3,00	23.141,62	14.528,30	6.370,04	31.854,71
Apple	iPhone 12	4,00	128,00	0,00	15,00	-14.183,89	31.863,55	-71.013,44	48.318,63
Apple	iPhone 12	4,00	128,00	2,00	4,00	-14.087,22	19.104,79	-42.584,01	-1.827,82

Apple	iPhone 12	4,00	256,00	0,00	3,00	10.905,97	68.466,48	-68.121,61	52.327,66
Apple	iPhone 12 Pro	6,00	128,00	0,00	11,00	5.863,08	34.842,21	-73.352,90	47.084,54
Apple	iPhone 12 Pro	6,00	256,00	0,00	7,00	-37.853,49	67.301,09	-121.016,11	40.611,13
Apple	iPhone 12 Pro Max	6,00	128,00	0,00	9,00	4.505,63	33.397,77	-63.549,87	53.639,25
Apple	iPhone 12 Pro Max	6,00	256,00	0,00	9,00	7.051,60	45.029,38	-71.454,59	59.161,28
Apple	iPhone 12 mini	4,00	64,00	0,00	5,00	19.333,60	24.672,72	-2.019,97	50.226,97
Apple	iPhone 12 mini	4,00	128,00	0,00	8,00	8.077,40	25.079,63	-37.235,67	33.568,63
Apple	iPhone 13	4,00	128,00	0,00	20,00	10.398,43	41.214,44	-103.841,01	74.509,99
Apple	iPhone 13	4,00	128,00	2,00	7,00	2.089,95	49.627,24	-101.693,95	52.545,33
Apple	iPhone 13	4,00	256,00	0,00	4,00	18.832,87	51.273,65	-36.618,56	77.679,96
Apple	iPhone 13	4,00	256,00	2,00	2,00	21.298,53	16.154,67	9.875,45	32.721,61
Apple	iPhone 13 Pro	6,00	128,00	0,00	10,00	-16.240,72	63.966,15	-117.831,92	70.471,13
Apple	iPhone 13 Pro	6,00	256,00	0,00	17,00	13.122,33	59.655,57	-103.587,93	84.836,19
Apple	iPhone 13 Pro Max	6,00	128,00	0,00	14,00	4.701,17	48.959,85	-81.315,12	85.073,76
Apple	iPhone 13 Pro Max	6,00	128,00	2,00	3,00	-76.116,47	40.095,72	-110.935,06	-32.279,62
Apple	iPhone 13 Pro Max	6,00	256,00	0,00	3,00	-47.642,02	20.484,69	-63.359,45	-24.475,00
Apple	iPhone 13 mini	4,00	128,00	0,00	6,00	-14.069,67	62.577,42	-87.020,63	82.948,78
Apple	iPhone 13 mini	4,00	128,00	2,00	4,00	-2.596,22	8.805,60	-11.392,76	5.035,41
Apple	iPhone 13 mini	4,00	256,00	2,00	3,00	-7.234,81	28.156,29	-38.961,05	14.781,46

Apple	iPhone 14	6,00	128,00	2,00	6,00	6.381,98	23.466,92	-25.338,48	42.547,09
Apple	iPhone 14	6,00	256,00	2,00	1,00	68.985,29		68.985,29	68.985,29
Apple	iPhone 14 Pro	6,00	256,00	2,00	1,00	8.428,55		8.428,55	8.428,55
Apple	iPhone 14 Pro Max	6,00	128,00	0,00	4,00	-34.990,32	37.532,33	-76.651,98	3.244,82
Apple	iPhone 14 Pro Max	6,00	128,00	2,00	1,00	2.941,78		2.941,78	2.941,78
Apple	iPhone 14 Pro Max	6,00	256,00	0,00	10,00	-38.499,61	132.823,68	-193.798,77	239.151,04
Apple	iPhone 14 Pro Max	6,00	256,00	2,00	3,00	-37.054,50	54.946,18	-84.954,89	22.926,90
Apple	iPhone 5s	1,00	16,00	0,00	4,00	1.071,83	6.887,18	-6.323,67	7.544,19
Apple	iPhone 6	1,00	16,00	0,00	10,00	-231,53	11.309,95	-16.981,71	18.989,50
Apple	iPhone 6	1,00	32,00	0,00	7,00	-3.224,50	8.067,40	-13.869,00	5.267,45
Apple	iPhone 6	1,00	64,00	0,00	12,00	-1.594,10	9.416,37	-18.601,49	12.109,12
Apple	iPhone 6 Plus	1,00	64,00	0,00	1,00	6.455,91		6.455,91	6.455,91
Apple	iPhone 6s	2,00	16,00	0,00	10,00	2.385,09	13.125,36	-17.368,29	23.329,99
Apple	iPhone 6s	2,00	32,00	0,00	8,00	-1.146,41	10.935,56	-18.597,46	15.388,80
Apple	iPhone 6s	2,00	64,00	0,00	17,00	-8.218,73	10.905,26	-26.269,15	10.794,75
Apple	iPhone 6s	2,00	128,00	0,00	1,00	14.609,73		14.609,73	14.609,73
Apple	iPhone 6s Plus	2,00	64,00	0,00	6,00	-9.779,40	7.803,72	-16.727,59	3.511,23
Apple	iPhone 7	2,00	32,00	0,00	50,00	-2.186,17	13.010,74	-32.610,04	22.585,33
Apple	iPhone 7	2,00	32,00	1,00	5,00	-5.290,81	18.734,27	-24.025,45	19.472,68
Apple	iPhone 7	2,00	128,00	0,00	38,00	2.497,14	15.723,43	-28.941,60	31.983,24
Apple	iPhone 7	2,00	256,00	0,00	4,00	-23.191,32	21.519,52	-49.086,00	2.886,16
Apple	iPhone 7 Plus	3,00	32,00	0,00	15,00	-421,85	22.113,49	-33.980,55	29.683,40
Apple	iPhone 7	3,00	128,00	0,00	36,00	-70,14	19.008,42	-40.859,19	43.214,49

	Plus								
Apple	iPhone 7 Plus	3,00	256,00	0,00	5,00	-3.825,08	17.203,46	-26.800,41	15.537,28
Apple	iPhone 8	2,00	64,00	0,00	60,00	4.903,54	17.211,32	-36.665,24	28.520,33
Apple	iPhone 8	2,00	64,00	1,00	3,00	-12.022,25	14.764,63	-25.029,16	4.026,33
Apple	iPhone 8	2,00	256,00	0,00	8,00	3.071,22	26.079,46	-51.677,43	26.450,91
Apple	iPhone 8 Plus	3,00	64,00	0,00	37,00	-3.779,28	21.261,39	-44.966,75	30.225,56
Apple	iPhone 8 Plus	3,00	256,00	0,00	15,00	-1.928,11	20.562,27	-41.086,35	35.727,51
Apple	iPhone SE (2020)	3,00	64,00	0,00	6,00	1.930,06	18.091,64	-26.686,54	25.131,88
Apple	iPhone SE (2020)	3,00	128,00	0,00	5,00	-7.939,02	34.479,59	-61.977,62	29.566,88
Apple	iPhone X	3,00	64,00	0,00	12,00	21.537,85	17.584,81	-11.211,31	41.916,17
Apple	iPhone X	3,00	256,00	0,00	20,00	14.691,42	21.573,07	-32.494,52	49.068,59
Apple	iPhone XS	4,00	64,00	0,00	12,00	15.228,14	31.653,71	-32.669,75	67.583,26
Apple	iPhone XS	4,00	256,00	0,00	7,00	-10.421,24	25.326,49	-43.959,62	22.068,21
Apple	iPhone XS Max	4,00	64,00	0,00	6,00	81,16	30.682,85	-41.062,56	46.051,50
Apple	iPhone XS Max	4,00	256,00	0,00	10,00	106,58	21.367,39	-32.693,84	31.962,18
Motorola	Edge 20 Pro	12,00	256,00	0,00	2,00	-2.744,05	14.634,88	-13.092,47	7.604,37
Motorola	Edge 30 Fusion	12,00	256,00	2,00	5,00	-2.311,16	13.556,69	-13.572,28	17.071,65
Motorola	Edge 30 Neo	8,00	128,00	2,00	13,00	4.304,43	4.415,39	-2.712,41	11.554,14
Motorola	Edge 30 Pro	12,00	256,00	0,00	4,00	5.821,13	16.461,07	-16.611,70	19.723,93
Motorola	Edge 30 Pro	12,00	256,00	2,00	3,00	-25.600,13	10.148,02	-36.274,43	-16.076,48

Motorola	Edge 30 Ultra	12,00	256,00	2,00	4,00	19.201,71	17.380,49	-6.549,55	30.891,65
Motorola	Moto G 30	4,00	128,00	0,00	2,00	4.380,86	5.768,13	302,17	8.459,54
Motorola	Moto G 51	4,00	128,00	2,00	5,00	-3.556,65	14.900,51	-30.015,46	5.697,56
Motorola	Moto G 7	4,00	64,00	0,00	2,00	4.868,69	11.592,31	-3.328,31	13.065,69
Motorola	Moto G 7 Play	2,00	32,00	0,00	2,00	2.413,99	833,14	1.824,87	3.003,11
Motorola	Moto G 7 Plus	4,00	64,00	0,00	1,00	-4.623,42		-4.623,42	-4.623,42
Motorola	Moto G 72	6,00	128,00	2,00	8,00	-1.612,18	7.553,50	-13.586,08	9.249,00
Motorola	Moto G 8 Plus	4,00	64,00	0,00	1,00	1.457,60		1.457,60	1.457,60
Motorola	Moto G 82	6,00	128,00	1,00	1,00	7.061,34		7.061,34	7.061,34
Motorola	Moto G 82	6,00	128,00	2,00	1,00	-5.827,45		-5.827,45	-5.827,45
Motorola	Moto G 9 Plus	4,00	128,00	0,00	12,00	1.830,92	11.524,97	-18.616,73	25.709,14
Samsung	Galaxy A 03s	4,00	64,00	2,00	10,00	-1.235,34	3.885,88	-8.699,79	2.773,95
Samsung	Galaxy A 04s	4,00	128,00	2,00	18,00	3.512,22	4.680,51	-10.673,01	11.318,67
Samsung	Galaxy A 12	4,00	64,00	0,00	6,00	3.464,97	5.096,19	-3.577,97	9.151,72
Samsung	Galaxy A 12	4,00	128,00	0,00	2,00	4.699,25	5.579,57	753,89	8.644,60
Samsung	Galaxy A 21s	4,00	64,00	0,00	2,00	-47,79	168,92	-167,23	71,66
Samsung	Galaxy A 21s	4,00	128,00	0,00	2,00	-5.907,82	18.004,47	-18.638,91	6.823,26
Samsung	Galaxy A 24 4G	6,00	128,00	2,00	3,00	2.730,37	9.426,97	-4.931,48	13.257,54
Samsung	Galaxy A 30	3,00	32,00	0,00	5,00	-750,20	2.015,23	-2.413,63	2.206,54
Samsung	Galaxy A	3,00	32,00	1,00	1,00	-7.135,53		-7.135,53	-7.135,53

	30								
Samsung	Galaxy A 31	4,00	128,00	0,00	7,00	-1.721,97	9.299,51	-16.171,83	9.150,91
Samsung	Galaxy A 33	6,00	128,00	0,00	5,00	-3.650,89	9.858,15	-17.688,83	7.993,14
Samsung	Galaxy A 33	6,00	128,00	2,00	7,00	-91,40	13.164,93	-23.212,04	12.332,20
Samsung	Galaxy A 34	6,00	128,00	2,00	6,00	-488,49	14.597,82	-20.305,35	16.730,27
Samsung	Galaxy A 50	4,00	64,00	0,00	4,00	4.259,97	8.405,04	-5.433,66	12.287,43
Samsung	Galaxy A 50	4,00	64,00	1,00	3,00	-2.565,49	929,82	-3.606,77	-1.818,20
Samsung	Galaxy A 50	4,00	128,00	0,00	2,00	-7.802,33	2.238,31	-9.385,05	-6.219,61
Samsung	Galaxy A 53 5G	6,00	128,00	0,00	3,00	14.561,10	3.937,51	11.331,85	18.947,54
Samsung	Galaxy A 53 5G	6,00	128,00	2,00	4,00	4.072,42	12.982,08	-7.870,37	22.551,63
Samsung	Galaxy A 54	8,00	128,00	2,00	7,00	-552,92	6.418,16	-11.106,50	4.687,47
Samsung	Galaxy A 54	8,00	256,00	2,00	8,00	-12.938,66	18.531,62	-35.435,56	10.591,73
Samsung	Galaxy A 70	6,00	128,00	0,00	2,00	3.732,15	11.013,50	-4.055,57	11.519,87
Samsung	Galaxy A 71	6,00	128,00	0,00	4,00	-9.519,51	8.404,63	-20.317,53	-1.019,56
Samsung	Galaxy S 10+	8,00	128,00	0,00	2,00	-1.224,77	6.125,33	-5.556,03	3.106,49
Samsung	Galaxy S 10e	6,00	128,00	0,00	5,00	-2.741,73	10.439,83	-13.447,93	11.587,19
Samsung	Galaxy S 21 FE 5G	6,00	128,00	0,00	2,00	-13.069,29	2.792,69	-15.044,02	-11.094,56
Samsung	Galaxy S 21 Ultra	12,00	256,00	0,00	4,00	18.477,26	32.215,27	-23.533,71	54.559,51

	5G								
Samsung	Galaxy S 21 Ultra 5G	12,00	256,00	1,00	1,00	-4.120,49		-4.120,49	-4.120,49
Samsung	Galaxy S 21+ 5G	8,00	128,00	0,00	2,00	-15.289,62	29.388,57	-36.070,47	5.491,24
Samsung	Galaxy S 22 5G	8,00	128,00	0,00	3,00	-25.147,03	51.232,79	-81.975,25	17.504,83
Samsung	Galaxy S 22 5G	8,00	128,00	2,00	6,00	-13.526,50	24.320,05	-47.039,93	25.971,09
Samsung	Galaxy S 22 5G	8,00	256,00	0,00	5,00	-15.043,76	50.742,54	-66.989,83	40.317,91
Samsung	Galaxy S 22 Ultra 5G	12,00	256,00	0,00	12,00	-24.815,85	28.288,44	-84.009,97	13.524,78
Samsung	Galaxy S 22 Ultra 5G	12,00	256,00	1,00	4,00	-51.549,85	40.056,91	-95.072,03	-1.965,98
Samsung	Galaxy S 22+ 5G	8,00	256,00	0,00	3,00	1.890,38	45.526,70	-48.262,31	40.612,65
Samsung	Galaxy S 23	8,00	256,00	0,00	3,00	14.758,54	66.889,92	-34.616,66	90.883,75
Samsung	Galaxy S 23	8,00	256,00	2,00	4,00	-71.058,06	54.757,77	-107.942,66	8.138,65
Samsung	Galaxy S 23 Ultra	12,00	256,00	0,00	1,00	-7.392,65		-7.392,65	-7.392,65
Samsung	Galaxy S 7	4,00	32,00	0,00	4,00	11.388,61	8.677,20	689,22	20.060,30
Samsung	Galaxy S 7	4,00	32,00	1,00	2,00	501,32	555,72	108,36	894,27
Samsung	Galaxy S 8+	4,00	64,00	0,00	5,00	-9.916,65	17.409,37	-26.154,03	9.629,67
Samsung	Galaxy S 9	4,00	64,00	0,00	6,00	-1.357,66	8.546,30	-14.075,55	9.138,24
Samsung	Galaxy S 9	4,00	64,00	1,00	2,00	2.871,66	11.815,99	-5.483,51	11.226,82
Samsung	Galaxy S 9+	6,00	64,00	0,00	7,00	-5.556,64	10.234,82	-20.186,76	8.508,79

Xiaomi	Mi 11T Pro	8,00	256,00	2,00	1,00	-2.044,87		-2.044,87	-2.044,87
Xiaomi	Poco X4 Pro 5G	6,00	128,00	2,00	3,00	-5.160,40	9.969,91	-16.048,33	3.522,20
Xiaomi	Poco X4 Pro 5G	8,00	256,00	2,00	8,00	1.553,25	10.029,19	-14.004,05	15.248,07
Xiaomi	Poco X5 Pro	8,00	256,00	2,00	5,00	8.084,27	11.676,92	-6.713,52	23.817,55
Xiaomi	Redmi Note 10 Pro	6,00	128,00	2,00	2,00	5.180,44	2.907,18	3.124,75	7.236,13
Xiaomi	Redmi Note 11	4,00	128,00	2,00	6,00	-334,94	5.438,56	-5.677,16	6.560,43
Xiaomi	Redmi Note 7	4,00	64,00	0,00	3,00	199,20	5.829,64	-4.661,02	6.662,73

Tabla 2. Potencial ganancia porcentual por dispositivo, capacidad y condición.

Marca	Modelo	ram_num	memory_num	item_condition	pred_1	pred_30	%profit
Apple	iPhone 11	4,00	64,00	0,00	225.806,82	281.561,81	24,69
Apple	iPhone 11	4,00	128,00	0,00	245.590,24	287.753,99	17,17
Apple	iPhone 11	4,00	256,00	0,00	237.949,39	297.015,70	24,82
Apple	iPhone 11 Pro	4,00	64,00	0,00	275.866,74	307.043,13	11,30
Apple	iPhone 11 Pro	4,00	256,00	0,00	296.430,69	328.721,37	10,89
Apple	iPhone 11 Pro Max	4,00	64,00	0,00	324.896,44	344.390,22	6,00
Apple	iPhone 11 Pro Max	4,00	256,00	0,00	335.220,18	350.374,12	4,52
Apple	iPhone 12	4,00	64,00	0,00	333.436,11	403.201,05	20,92
Apple	iPhone 12	4,00	128,00	0,00	344.829,69	422.212,13	22,44
Apple	iPhone 12	4,00	256,00	0,00	340.222,56	431.758,55	26,90

Apple	iPhone 12 mini	4,00	64,00	0,00	272.277,53	321.220,84	17,98
Apple	iPhone 12 mini	4,00	128,00	0,00	284.301,72	336.245,69	18,27
Apple	iPhone 12 Pro	6,00	128,00	0,00	383.622,06	429.532,36	11,97
Apple	iPhone 12 Pro	6,00	256,00	0,00	415.464,78	444.768,11	7,05
Apple	iPhone 12 Pro Max	6,00	128,00	0,00	431.548,35	504.247,89	16,85
Apple	iPhone 12 Pro Max	6,00	256,00	0,00	441.803,41	513.581,53	16,25
Apple	iPhone 13	4,00	128,00	0,00	413.493,92	463.024,23	11,98
Apple	iPhone 13	4,00	256,00	0,00	415.066,11	473.198,24	14,01
Apple	iPhone 13 mini	4,00	128,00	0,00	388.499,98	422.066,79	8,64
Apple	iPhone 13 Pro	6,00	128,00	0,00	487.995,90	512.754,48	5,07
Apple	iPhone 13 Pro	6,00	256,00	0,00	518.785,58	532.114,39	2,57
Apple	iPhone 13 Pro Max	6,00	128,00	0,00	537.065,72	606.529,04	12,93
Apple	iPhone 13 Pro Max	6,00	256,00	0,00	515.402,35	620.700,30	20,43
Apple	iPhone 14 Pro Max	6,00	128,00	0,00	644.267,01	791.264,17	22,82
Apple	iPhone 14 Pro Max	6,00	256,00	0,00	623.470,22	811.074,33	30,09
Apple	iPhone 5s	1,00	16,00	0,00	17.886,68	18.944,90	5,92
Apple	iPhone 6	1,00	16,00	0,00	31.535,46	36.474,91	15,66
Apple	iPhone 6	1,00	32,00	0,00	36.050,63	42.295,93	17,32
Apple	iPhone 6	1,00	64,00	0,00	34.151,85	39.947,09	16,97
Apple	iPhone 6	1,00	64,00	0,00	49.862,47	56.508,00	13,33

	Plus						
Apple	iPhone 6s	2,00	16,00	0,00	37.648,78	38.141,16	1,31
Apple	iPhone 6s	2,00	32,00	0,00	43.023,91	44.242,41	2,83
Apple	iPhone 6s	2,00	64,00	0,00	44.249,43	46.764,39	5,68
Apple	iPhone 6s	2,00	128,00	0,00	49.308,22	49.121,39	-0,38
Apple	iPhone 6s Plus	2,00	64,00	0,00	60.578,23	65.863,24	8,72
Apple	iPhone 7	2,00	32,00	0,00	58.666,40	84.473,86	43,99
Apple	iPhone 7	2,00	128,00	0,00	68.504,04	87.471,07	27,69
Apple	iPhone 7	2,00	256,00	0,00	71.547,93	93.373,09	30,50
Apple	iPhone 7 Plus	3,00	32,00	0,00	86.494,32	107.429,17	24,20
Apple	iPhone 7 Plus	3,00	128,00	0,00	94.297,83	111.282,46	18,01
Apple	iPhone 7 Plus	3,00	256,00	0,00	103.428,90	119.016,92	15,07
Apple	iPhone 8	2,00	64,00	0,00	95.864,86	112.289,39	17,13
Apple	iPhone 8	2,00	256,00	0,00	105.349,92	116.958,82	11,02
Apple	iPhone 8 Plus	3,00	64,00	0,00	121.727,64	128.956,75	5,94
Apple	iPhone 8 Plus	3,00	256,00	0,00	135.121,63	145.609,40	7,76
Apple	iPhone SE (2020)	3,00	64,00	0,00	133.168,90	149.719,32	12,43
Apple	iPhone SE (2020)	3,00	128,00	0,00	132.671,36	143.237,41	7,96
Apple	iPhone X	3,00	64,00	0,00	159.227,56	158.337,09	-0,56
Apple	iPhone X	3,00	256,00	0,00	154.881,13	180.572,01	16,59
Apple	iPhone XS	4,00	64,00	0,00	157.969,12	161.269,40	2,09
Apple	iPhone XS	4,00	256,00	0,00	188.140,11	202.265,89	7,51
Apple	iPhone XS	4,00	64,00	0,00	180.713,47	194.811,26	7,80

	Max						
Apple	iPhone XS Max	4,00	256,00	0,00	183.152,28	200.258,66	9,34
Motorola	Edge 20 Pro	12,00	256,00	0,00	153.731,67	186.333,99	21,21
Motorola	Edge 30 Pro	12,00	256,00	0,00	206.089,27	268.666,78	30,36
Motorola	Moto G 30	4,00	128,00	0,00	62.850,00	76.202,27	21,24
Motorola	Moto G 7	4,00	64,00	0,00	49.323,69	56.450,77	14,45
Motorola	Moto G 7 Play	2,00	32,00	0,00	31.138,74	33.971,79	9,10
Motorola	Moto G 7 Plus	4,00	64,00	0,00	47.632,39	54.023,55	13,42
Motorola	Moto G 8 Plus	4,00	64,00	0,00	55.808,21	68.388,01	22,54
Motorola	Moto G 9 Plus	4,00	128,00	0,00	69.704,29	88.271,70	26,64
Samsung	Galaxy A 12	4,00	64,00	0,00	54.350,00	68.031,29	25,17
Samsung	Galaxy A 12	4,00	128,00	0,00	57.680,33	69.657,29	20,76
Samsung	Galaxy A 21s	4,00	64,00	0,00	64.463,59	74.151,75	15,03
Samsung	Galaxy A 21s	4,00	128,00	0,00	63.639,67	73.956,88	16,21
Samsung	Galaxy A 30	3,00	32,00	0,00	44.243,86	50.646,98	14,47
Samsung	Galaxy A 31	4,00	128,00	0,00	60.652,20	73.647,79	21,43
Samsung	Galaxy A 33	6,00	128,00	0,00	107.735,21	125.910,29	16,87
Samsung	Galaxy A 50	4,00	64,00	0,00	55.162,05	64.579,24	17,07
Samsung	Galaxy A	4,00	128,00	0,00	59.530,09	67.351,45	13,14

	50						
Samsung	Galaxy A 53 5G	6,00	128,00	0,00	145.363,87	163.368,81	12,39
Samsung	Galaxy A 70	6,00	128,00	0,00	80.375,37	86.636,62	7,79
Samsung	Galaxy A 71	6,00	128,00	0,00	96.644,92	108.306,97	12,07
Samsung	Galaxy S 10+	8,00	128,00	0,00	135.925,42	137.255,73	0,98
Samsung	Galaxy S 10e	6,00	128,00	0,00	95.227,45	107.255,72	12,63
Samsung	Galaxy S 21+ 5G	8,00	128,00	0,00	203.927,08	204.085,02	0,08
Samsung	Galaxy S 21 FE 5G	6,00	128,00	0,00	204.211,01	204.830,36	0,30
Samsung	Galaxy S 21 Ultra 5G	12,00	256,00	0,00	273.827,66	304.880,92	11,34
Samsung	Galaxy S 22+ 5G	8,00	256,00	0,00	334.212,14	328.012,60	-1,85
Samsung	Galaxy S 22 5G	8,00	128,00	0,00	254.147,73	307.341,98	20,93
Samsung	Galaxy S 22 5G	8,00	256,00	0,00	352.892,01	349.718,25	-0,90
Samsung	Galaxy S 22 Ultra 5G	12,00	256,00	0,00	376.286,59	369.311,92	-1,85
Samsung	Galaxy S 23	8,00	256,00	0,00	414.185,53	429.618,63	3,73
Samsung	Galaxy S 23 Ultra	12,00	256,00	0,00	486.619,00	482.459,48	-0,85
Samsung	Galaxy S 7	4,00	32,00	0,00	42.818,65	50.899,92	18,87
Samsung	Galaxy S 8+	4,00	64,00	0,00	42.335,60	52.441,62	23,87
Samsung	Galaxy S	6,00	64,00	0,00	76.536,69	90.791,31	18,62

	9+						
Samsung	Galaxy S 9	4,00	64,00	0,00	65.613,52	82.819,84	26,22
Xiaomi	Redmi Note 7	4,00	64,00	0,00	58.957,28	66.313,56	12,48
Apple	iPhone 11	4,00	64,00	1,00	342.836,23	386.602,45	12,77
Apple	iPhone 11	4,00	128,00	1,00	342.068,07	374.170,23	9,38
Apple	iPhone 11 Pro	4,00	64,00	1,00	377.840,60	403.792,16	6,87
Apple	iPhone 12	4,00	64,00	1,00	395.966,35	478.913,00	20,95
Apple	iPhone 7	2,00	32,00	1,00	87.899,57	109.774,26	24,89
Apple	iPhone 8	2,00	64,00	1,00	135.299,07	160.176,71	18,39
Motorola	Moto G 82	6,00	128,00	1,00	155.448,87	160.057,71	2,96
Samsung	Galaxy A 30	3,00	32,00	1,00	50.124,06	56.726,93	13,17
Samsung	Galaxy A 50	4,00	64,00	1,00	64.897,24	70.486,02	8,61
Samsung	Galaxy S 21 Ultra 5G	12,00	256,00	1,00	290.702,61	323.708,70	11,35
Samsung	Galaxy S 22 Ultra 5G	12,00	256,00	1,00	350.594,79	351.616,16	0,29
Samsung	Galaxy S 7	4,00	32,00	1,00	51.816,55	60.271,30	16,32
Samsung	Galaxy S 9	4,00	64,00	1,00	74.223,10	92.466,78	24,58
Apple	iPhone 11	4,00	64,00	2,00	340.327,97	376.411,70	10,60
Apple	iPhone 11	4,00	128,00	2,00	350.424,49	374.635,14	6,91
Apple	iPhone 12	4,00	64,00	2,00	468.347,58	511.335,83	9,18
Apple	iPhone 12	4,00	128,00	2,00	486.286,79	542.922,04	11,65
Apple	iPhone 13	4,00	128,00	2,00	558.831,24	587.950,30	5,21
Apple	iPhone 13	4,00	256,00	2,00	603.044,14	641.216,03	6,33
Apple	iPhone 13	4,00	128,00	2,00	516.222,27	575.272,31	11,44

	mini						
Apple	iPhone 13 mini	4,00	256,00	2,00	585.080,03	636.148,07	8,73
Apple	iPhone 13 Pro Max	6,00	128,00	2,00	813.343,75	873.499,62	7,40
Apple	iPhone 14	6,00	128,00	2,00	595.505,19	657.223,36	10,36
Apple	iPhone 14	6,00	256,00	2,00	711.126,82	788.392,45	10,87
Apple	iPhone 14 Pro	6,00	256,00	2,00	879.458,82	929.789,10	5,72
Apple	iPhone 14 Pro Max	6,00	128,00	2,00	899.553,26	942.856,41	4,81
Apple	iPhone 14 Pro Max	6,00	256,00	2,00	870.626,59	951.055,80	9,24
Motorola	Edge 30 Fusion	12,00	256,00	2,00	247.820,84	261.319,30	5,45
Motorola	Edge 30 Neo	8,00	128,00	2,00	192.301,46	188.148,11	-2,16
Motorola	Edge 30 Pro	12,00	256,00	2,00	292.111,43	360.243,88	23,32
Motorola	Edge 30 Ultra	12,00	256,00	2,00	393.366,19	417.599,94	6,16
Motorola	Moto G 51	4,00	128,00	2,00	129.137,48	146.528,12	13,47
Motorola	Moto G 72	6,00	128,00	2,00	136.366,06	142.930,48	4,81
Motorola	Moto G 82	6,00	128,00	2,00	156.295,19	161.828,98	3,54
Samsung	Galaxy A 03s	4,00	64,00	2,00	73.062,74	79.887,04	9,34
Samsung	Galaxy A 04s	4,00	128,00	2,00	88.473,92	99.464,19	12,42
Samsung	Galaxy A 24 4G	6,00	128,00	2,00	126.970,71	137.579,82	8,36
Samsung	Galaxy A 33	6,00	128,00	2,00	142.669,21	166.085,81	16,41
Samsung	Galaxy A	6,00	128,00	2,00	175.156,26	180.774,39	3,21

	34						
Samsung	Galaxy A 53 5G	6,00	128,00	2,00	192.721,02	225.913,66	17,22
Samsung	Galaxy A 54	8,00	128,00	2,00	199.233,66	214.988,16	7,91
Samsung	Galaxy A 54	8,00	256,00	2,00	211.533,46	238.901,38	12,94
Samsung	Galaxy S 22 5G	8,00	128,00	2,00	301.558,76	360.771,94	19,64
Samsung	Galaxy S 23	8,00	256,00	2,00	443.602,01	446.205,82	0,59
Xiaomi	Mi 11T Pro	8,00	256,00	2,00	327.044,17	365.953,88	11,90
Xiaomi	Poco X4 Pro 5G	6,00	128,00	2,00	212.560,35	230.568,57	8,47
Xiaomi	Poco X4 Pro 5G	8,00	256,00	2,00	246.631,86	278.331,46	12,85
Xiaomi	Poco X5 Pro	8,00	256,00	2,00	265.308,44	276.428,01	4,19
Xiaomi	Redmi Note 10 Pro	6,00	128,00	2,00	195.846,33	195.454,30	-0,20
Xiaomi	Redmi Note 11	4,00	128,00	2,00	107.497,49	126.191,11	17,39

Tabla 3. Potencial ganancia bruta porcentual por reacondicionamiento y espera.

Marca	Modelo	ram_num	memory_num	pred_1	pred_30	%profit
Apple	iPhone 11	4,00	64,00	225.806,82	386.602,45	71,21
Apple	iPhone 11	4,00	128,00	245.590,24	374.170,23	52,36
Apple	iPhone 11 Pro	4,00	64,00	275.866,74	403.792,16	46,37
Apple	iPhone 12	4,00	64,00	333.436,11	478.913,00	43,63

Apple	iPhone 7	2,00	32,00	58.666,40	109.774,26	87,12
Apple	iPhone 8	2,00	64,00	95.864,86	160.176,71	67,09
Samsung	Galaxy A 30	3,00	32,00	44.243,86	56.726,93	28,21
Samsung	Galaxy A 50	4,00	64,00	55.162,05	70.486,02	27,78
Samsung	Galaxy S 21 Ultra 5G	12,00	256,00	273.827,66	323.708,70	18,22
Samsung	Galaxy S 22 Ultra 5G	12,00	256,00	376.286,59	351.616,16	-6,56
Samsung	Galaxy S 7	4,00	32,00	42.818,65	60.271,30	40,76
Samsung	Galaxy S 9	4,00	64,00	65.613,52	92.466,78	40,93