

# Predicción del nivel de aceptación de un producto de una E-commerce (Flipkart) con implementación de Machine Learning



# Índice

1. Interés del estudio
2. Objetivos
3. Solución propuesta
4. Resultados
5. Conclusiones y trabajo futuro



# 1. Interés del estudio



# 1. Interés del estudio



## Optimización de Inventarios

Gestionar de manera más eficaz sus inventarios, asegurando que los productos más populares estén disponibles



01

## Estrategia de Marketing y Promoción

Los insights obtenidos a través de este estudio permiten a la empresa diseñar estrategias de marketing y promoción más informadas



02

## Satisfacción del Cliente

Manteniendo un stock adecuado de productos populares y promocionando los productos mejorando la experiencia del cliente



03

## Reducción de Costos

La optimización de inventarios y estrategias de marketing basadas en datos contribuyen a la reducción de costos



04

## Toma de Decisiones

Fomenta una cultura basada en toma de decisiones basada en datos



08

## Innovación

Oportunidad de avanzar en el campo de la analítica avanzada y machine learning



07

## Desarrollo de Productos

Los patrones identificados en la predicción del éxito de los productos pueden ser retroalimentados al proceso de Desarrollo de nuevos productos



06

## Competitividad y Cuota de Mercado

La habilidad de predecir el éxito de los productos y ajustar operaciones son factores que contribuyen a la ventaja competitiva



05



## 2. Objetivos



## 2. Objetivos

### PROBLEMA

Cliente desea comprar productos con FlipKart pero desconoce si el producto elegido será aceptado por los compradores finales.

### SOLUCIÓN

Desarrollar un modelo de machine learning que, utilizando los datos históricos de compras de los productos en la tienda e-commerce Flipkart, permita predecir el grado de aceptación de un producto.

Este grado de aceptación se determinará a partir de variables como número de ventas, reseñas de los clientes y frecuencia de devoluciones.

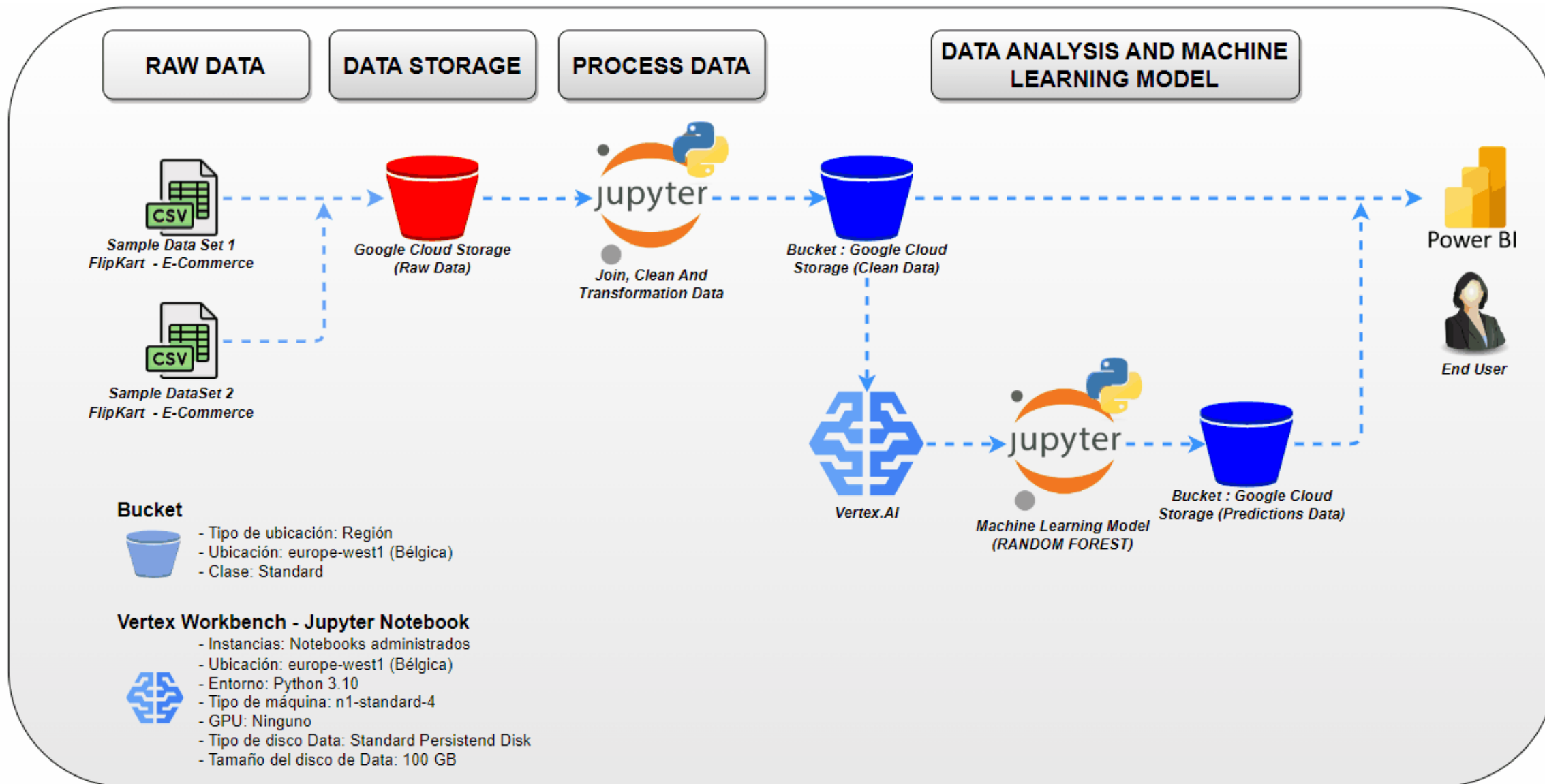
El objetivo final es anticipar cómo un producto será recibido por la comunidad de compradores y tomar decisiones informadas sobre estrategias de marketing, inventario y precios.



## 3. Solución Propuesta

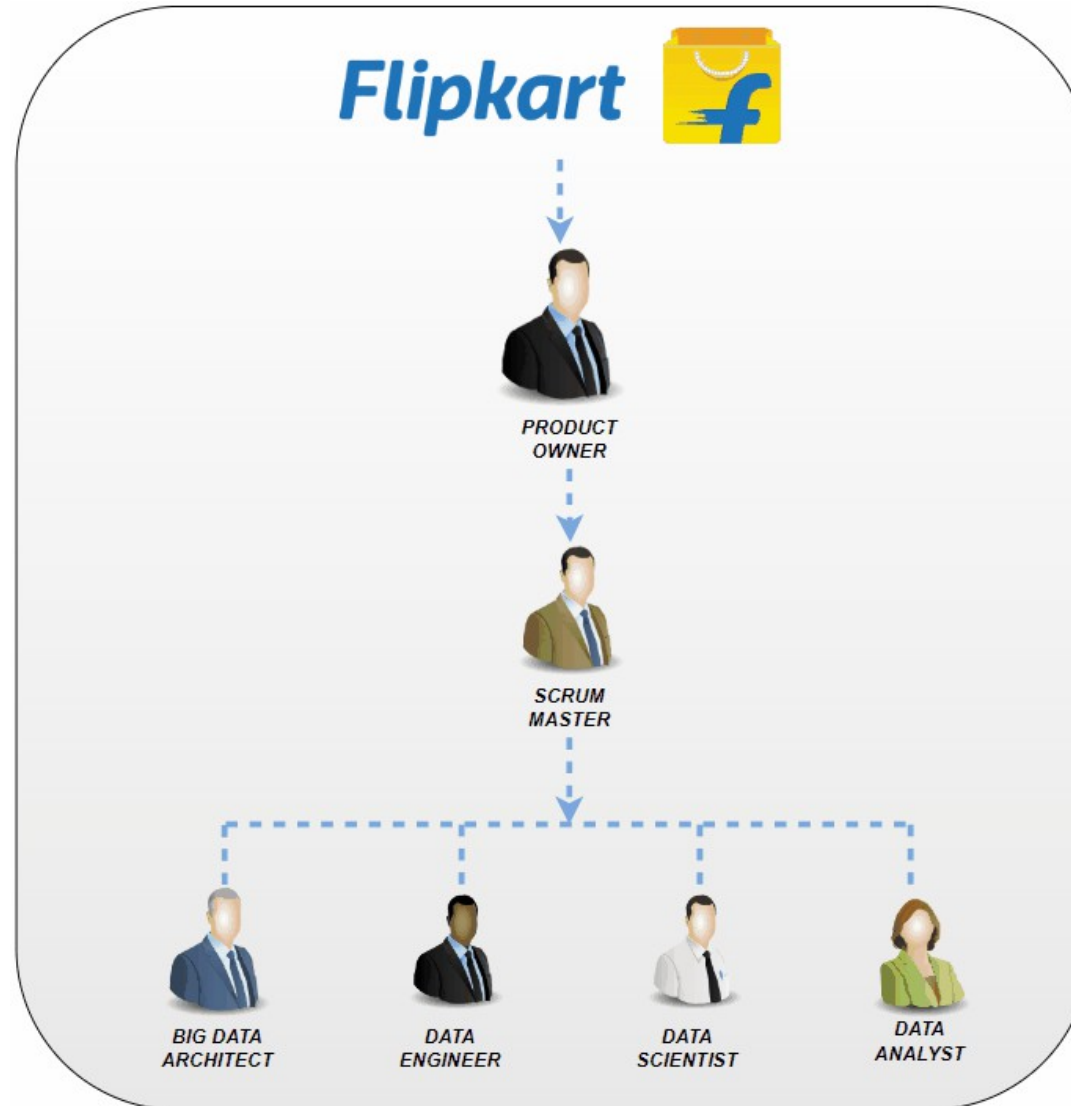


## 3. Solución Propuesta





### 3. Solución Propuesta



## 4. Resultados



RESULTADOS	PRICING ML	CLASSIFICATION ML
Características Iniciales	'brand' y 'sub_category' eran fundamentales para determinar el precio de un producto.	Las características 'brand', 'category', 'sub_category' y 'title' son esenciales
Nuevas Características Relevantes después de Aplicar El ML	'seller' y 'product_details' también eran importantes	La característica 'seller' también era crucial para determinar la aceptación de los productos
Confirmación de Métricas a través de Matriz de Correlación:	Matriz de Correlación ayudó a visualizar y confirmar la relación entre 'seller' y 'brand', validando así las métricas seleccionadas.	Matriz de Correlación corrobora así la adecuación de las métricas seleccionadas.
Evaluación de Modelos y Precisión:	Se evaluaron tres modelos: Regresión Lineal ( $R^2$ train: 86.8%, $R^2$ test: 86.7%), Árbol de Decisión (Precisión en entrenamiento: 94.5%, Precisión en prueba: 91.7%), y Random Forest (Precisión en entrenamiento: 79.2%, Precisión en prueba: 77.9%).	Se evaluaron dos modelos: XGBoost (Precisión en entrenamiento: 82%, Precisión en prueba: 78%) y Random Forest (Precisión en entrenamiento: 90%, Precisión en prueba: 83%).



RESULTADOS	PRICING ML	CLASSIFICATION ML
Selección de Modelo	Se seleccionó el modelo Random Forest por su capacidad de generalización en la predicción de precios y su robustez ante valores atípicos, lo que lo hace adaptable a diferentes datasets.	Se optó por el modelo Random Forest debido a un problema con el rango de etiquetado en XGBoost, que requiere etiquetas de 0 a 4 en lugar de 1 a 5. Aunque XGBoost generalizaba mejor, la flexibilidad de Random Forest ante esta particularidad de etiquetado lo hizo la elección preferida para evitar errores en caso de ausencia de registros para alguna etiqueta en particular.



## 4. Resultados

### The Choice



Coat	Dhoti	Jacket	Jeans	Kurta	Pants	Shirt	Shorts	Trousers	T-shirt
------	-------	--------	-------	-------	-------	-------	--------	----------	---------

Brand	Predicted Price
Adam Park	2034,50
Asa	1687,80
Bra	1534,00
Breakboun	1630,60
Byford by Pantaloo	1440,00
chawla fashi	1615,00
Crocks Cl	2674,25
DiscountZila Fashi	2317,00
G	3202,00
Inspi	1875,50
Lev	2354,50
MagMatri	1681,00
mark morn	1875,00
<b>Total</b>	<b>2106,49</b>

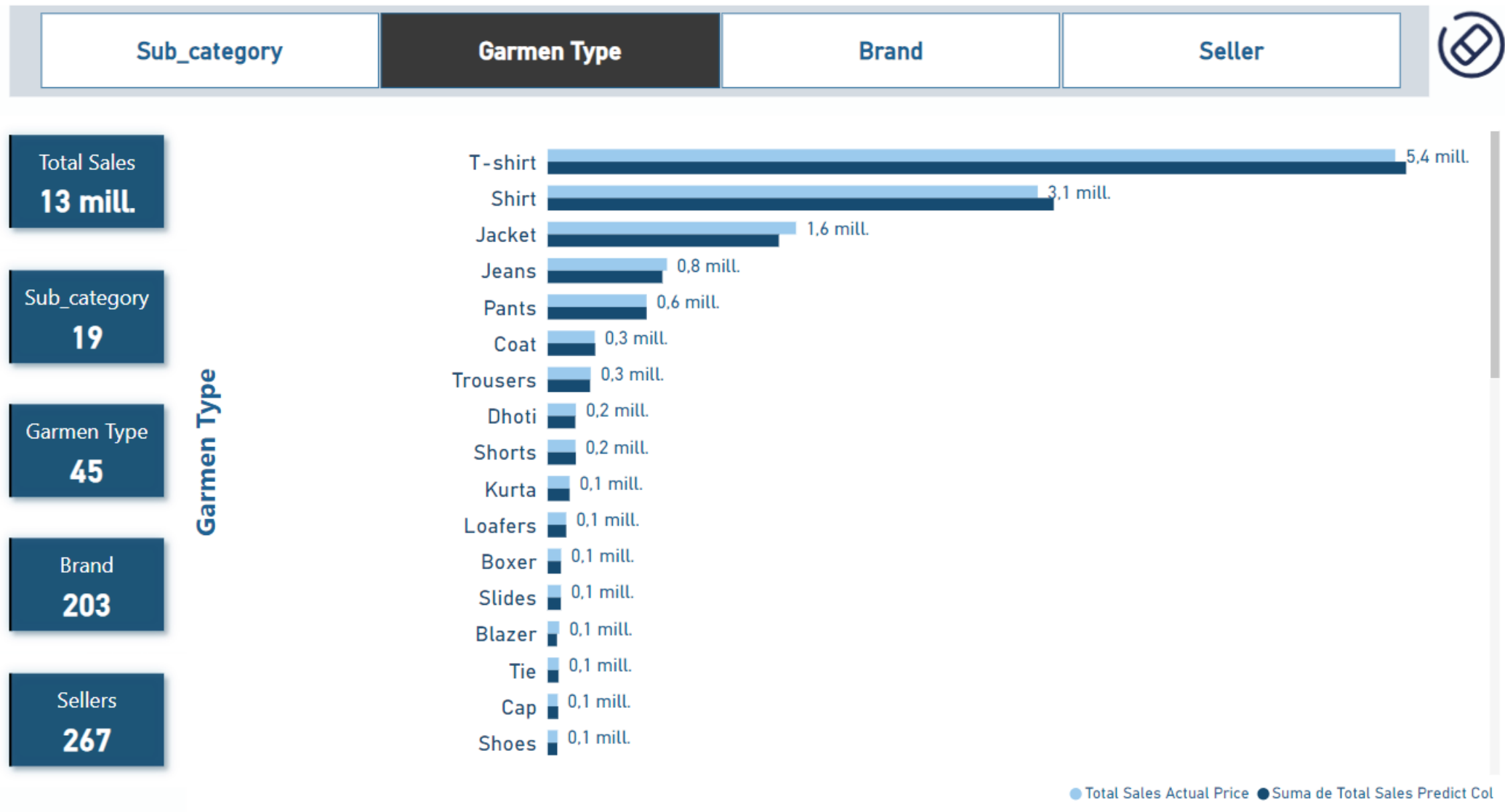
Brand	Predicted Rating
Adam Park	5,00
Asa	4,00
Bra	4,50
Breakboun	4,00
Byford by Pantaloo	5,00
chawla fashi	3,00
Crocks Cl	4,00
DiscountZila Fashi	4,00
FABUNIFOR	4,00
Fesn	3,00
INSPI	4,06
K	3,00
Lafant	3,00
<b>Total</b>	<b>3,87</b>

Seller	Predicted Rating
Unleashed clothin	4,00
UNKNOWN	4,17
UberUrban	3,79
SRCP	3,50
SPARKLING SYNTEX	3,50
SouthIndiaShoppin gMall	5,00
SMART SHOPPING123	3,00
RUPALFASHI	3,60
RetailNet	5,00
priyayarn	5,00
PLATINUM retails	4,00
MountainColours	5,00
<b>Total</b>	<b>3,87</b>



# 4. Resultados

## Price - Prediction



## 5. Conclusiones y Trabajo Futuro



## 5. Conclusiones y Trabajos Futuros

# Conclusiones

### Identificación de Características Relevantes:

El estudio permitió identificar características adicionales como 'seller', que resultaron ser relevantes para determinar la aceptación de los productos, extendiendo la comprensión inicial que se tenía sobre las variables influyentes.

### Validación de Métricas:

La Matriz de Correlación demostró ser una herramienta valiosa para validar las métricas seleccionadas y visualizar las relaciones entre diferentes características.

### Efectividad del Modelo Random Forest:

A pesar de ciertas limitaciones, el modelo Random Forest demostró ser una elección robusta y flexible para el problema en cuestión, especialmente en lo que respecta al manejo de etiquetas.

### Adecuación de la Arquitectura:

La arquitectura seleccionada resultó ser adecuada para el tipo de datos (batch) y el tamaño de los datasets, cumpliendo con los requerimientos sin necesidad de una gran escalabilidad.

### Potencial de Mejora Identificado:

El estudio reveló que hay potencial para mejorar la predicción incorporando técnicas avanzadas como Deep Learning o NLP, especialmente considerando la disponibilidad de datos de imágenes y reviews.

# Trabajos Futuros

### Incorporación de Deep Learning:

Explotar la información contenida en las imágenes de los productos mediante técnicas de Deep Learning puede proporcionar insights valiosos sobre cómo la apariencia de un producto impacta en su aceptación.

### Aplicación de Procesamiento de Lenguaje Natural (NLP):

Utilizar NLP para analizar las reviews de los productos puede ayudar a entender mejor las opiniones de los clientes y cómo estas afectan la aceptación del producto.

### Optimización del Etiquetado:

Revisar y mejorar el proceso de etiquetado podría evitar problemas como el encontrado en XGBoost, permitiendo explorar más modelos y técnicas.

### Exploración de Más Modelos y Técnicas:

Explorar otros modelos y técnicas de ML avanzadas puede proporcionar mejoras en la precisión de la predicción y ofrecer una comprensión más profunda de las variables que afectan la aceptación del producto.

### Ampliación del Dataset:

Ampliar el dataset con más datos históricos o más características podría enriquecer el modelo y mejorar la precisión de las predicciones, además de permitir la exploración de modelos más complejos como los basados en Deep Learning o NLP.





# Muchas gracias

**Dayana Franco**  
**Wilbert Vong**  
**Marcos Salafranca**  
**Jefferson Osorio**

