
Data Science Challenge



**mercado
libre**

Maria Camila Garzón Hortúa • 20.12.2023

Resumen

1. **Análisis de Datos**

- Exploración Inicial
- Hipótesis
- Sigüientes pasos

2. **Similitud de Productos**

- Modelos de lenguaje
- Puntaje de similitud

3. **Predicción de Fallas**

- Análisis previo
 - Modelos predictivos
 - Conclusiones
-

EDA

Ofertas

Relámpago



Exploración Inicial

- Relación matemática entre las variables, por tanto están muy **correlacionadas** entre ellas.
 - Sold_Amount y Sold_Quantity tienen una correlación muy alta y **hacen falta** en los mismos registros
 - El tipo de origen no representa información relevante al análisis
 - Más del 45% de las ofertas no tuvieron venta
-

Ofertas Relámpago en Mercado Libre

Hallazgos

- Relación entre las ventas y la fecha
- Verticales con mucha oferta y poca venta
- Las ofertas se interrumpen al acabar con el stock

Recomendaciones

- Momentos de lanzar ofertas
 - Tipos de ofertas
 - Ofertas sin stock
-

Fechas

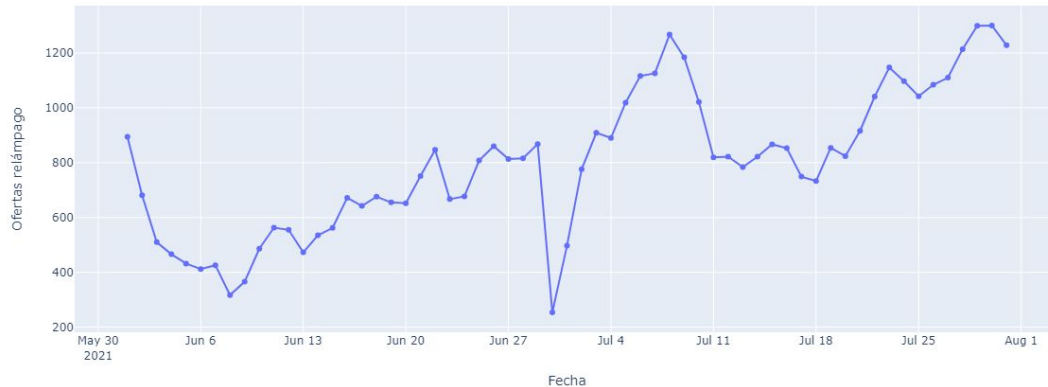
Información desde **Junio 1 hasta 31 Julio de 2021.**

Contexto global: **Pandemia**

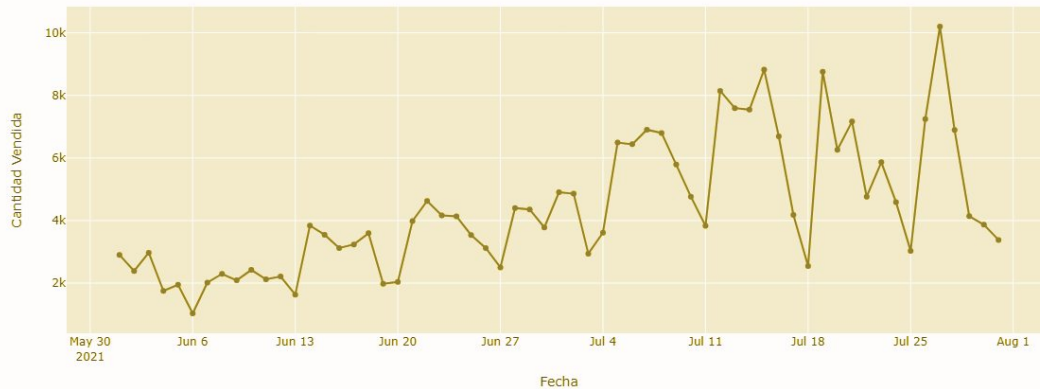
Contexto México: **Verano**



Línea de Tiempo vs Numero de ofertas relámpago



Línea de Tiempo vs Cantidad Vendida

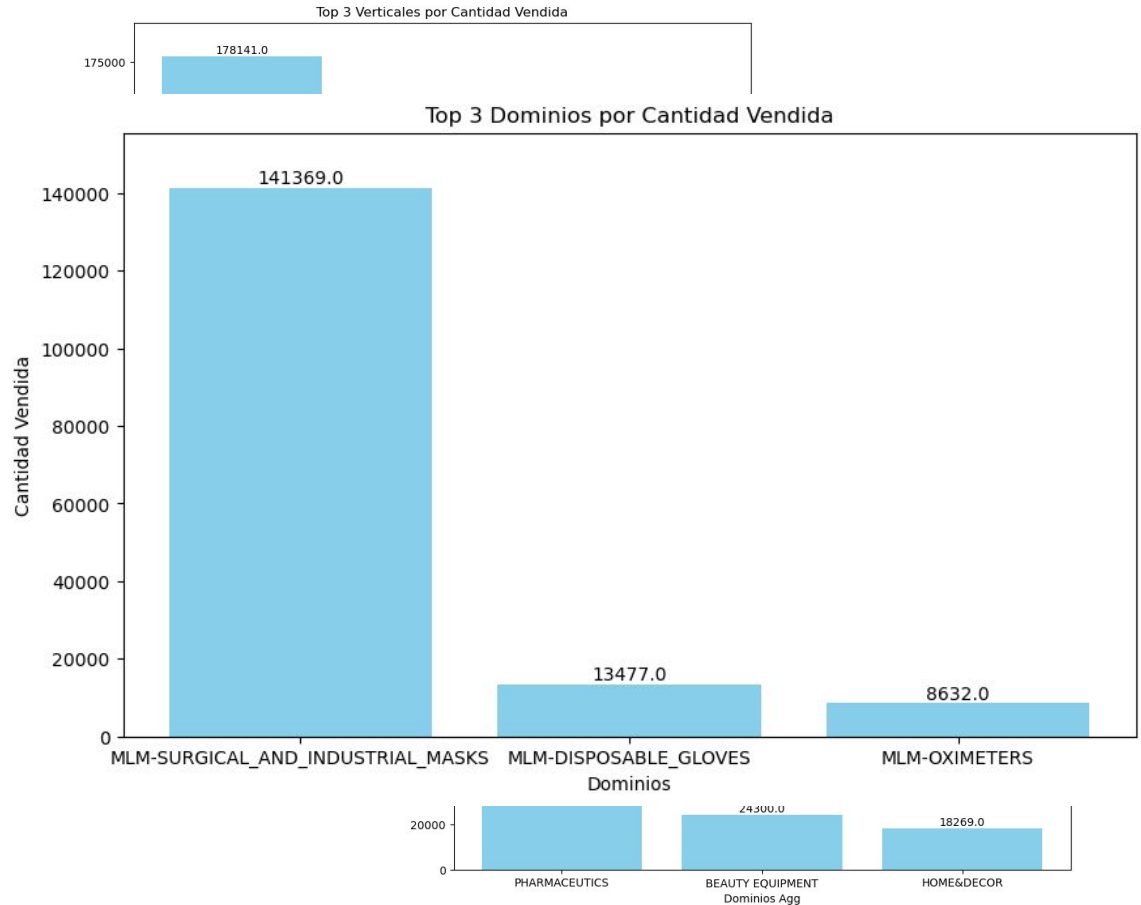


Categorías

¿Realmente cuáles son las que generan mayor número de ventas?

Contexto global: **Pandemia**

Contexto México: **Verano**



Stock

¿Qué sucede con las ventas de las unidades negativas?



2,6%

¿Por qué no se ha generado venta en esas ofertas?

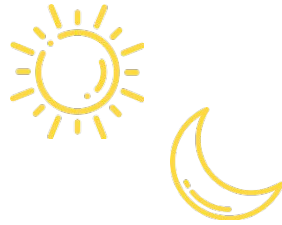


49%

49%

De las ofertas realizadas, no generan ninguna venta

Siguientes Pasos



Analizar la hora
o momento del
día con las
ventas



Obtener el % de
descuento
ofrecido



Construir
nuevas variables
para eliminar
correlaciones

Precio producto

% ventas

Recomendaciones



Políticas para
elegibilidad de
productos en
ofertas
relámpago

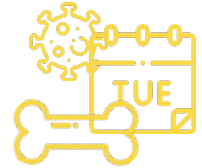


Promover las
ofertas en fines
de semana

Recomendar a
los vendedores
el **mejor día** para
poner su oferta



Alertar a los
vendedores
cuando quede
poco inventario
para que no se
cierre la oferta



Realizar ofertas
específicas
según el **día de la
semana** o
temporada.

Ej: Martes de
Mascotas








Similitud de Productos



Comprador

- Comparación de precios
- Comparar tiempos de entrega
- Recomendar productos sustitutos

¿Cuáles son esos productos a comparar?

<div><div><p>Lampara De Escritorio Led, Lampara De Escritorio De Brazo</p><p>(Este producto)</p></div><div><p>Lampara Para Portatil O Monitor Para Iluminar Teclado Grande Color De...</p><p>Ver producto</p></div><div><p>Lámpara Led Escritorio Cargador Inalámbrico Y Temporizador</p><p>Ver producto</p></div><div><p>Pack Lámparas De Mesa 40 W 2 Unidades</p><p>Ver producto</p></div></div>				
Opiniones	—	4.2  (9)	5.0  (2)	4.7  (15)
Precio	\$ 119.147	\$ 100.137	\$ 150.900	\$ 119.900
Con USB	Si	Si	—	—
Color de la estructura	—	Negro	—	Plata
Color de la pantalla	—	Blanco	—	—
Volaje	—	9V	—	—
Ancho total	—	50 cm	11.99 cm	12.5 cm
Tecnología de iluminación	—	LED	—	—
Material de la estructura	—	Plastico	Plastico	Metal
Altura total	—	—	39.88 cm	30 cm

Embeddings

Representaciones vectoriales de las palabras o frases.

- Análisis sintáctico
- Análisis de sentimiento
- Similitud entre texto



ITE_ITEM_TITLE

Tênis Feminino Infantil Cano Alto Estrela Coração Glitter
Tênis Masculino Sport Academia/caminhada Esporte Azul
Sapatilhas Sapatos Esportivos Sapatos Casuais Homem Running
Tenis Ous Emergente Copper Essencial
1879202 Tênis Ramarim De Amarrar
Sapatênis Couro Legítimo Forrado Em Couro Solado Costurado
Tenis Feminino Casual Chunky Vizzano Dad Sneaker Tratorado
Tênis Fila D-formation - Feminino
Tênis Infantil Pé Com Pé Masculino Calce Fácil Azul Menino
Tênis Feminino Luana Têxtil Malu Super Comfort Azul Marinho

Modelos de lenguaje

Sentence Transformers

- Redes siamesas
- Comparación con Similitud del Coseno
- Modelo multilingual

<https://www.sbert.net/docs/training/overview.html>

Universal Sentence encoder




- Desarrollado por Google
- Comparación con correlación

<https://static.googleusercontent.com/media/research.google.com/en/pubs/archive/46808.pdf>

Selección de Modelo Multilingual

Model Name	Performance	Performance	Avg.	Speed	Model Size
	Sentence Embeddings (14 Datasets) ⓘ	Semantic Search (6 Datasets) ⓘ	Performance ⓘ		
all-mpnet-base-v2 ⓘ	69.57	57.02	63.30	2800	420 MB
multi-qa-mpnet-base-dot-v1 ⓘ	66.76	57.60	62.18	2800	420 MB
all-distilroberta-v1 ⓘ	68.73	50.94	59.84	4000	290 MB
all-MiniLM-L12-v2 ⓘ	68.70	50.82	59.76	7500	120 MB
multi-qa-distilbert-cos-v1 ⓘ	65.98	52.83	59.41	4000	250 MB
all-MiniLM-L6-v2 ⓘ	68.06	49.54	58.80	14200	80 MB
multi-qa-MiniLM-L6-cos-v1 ⓘ	64.33	51.83	58.08	14200	80 MB
paraphrase-multilingual-mpnet-base-v2 ⓘ	65.83	41.68	53.75	2500	970 MB
paraphrase-albert-small-v2 ⓘ	64.46	40.04	52.25	5000	43 MB
paraphrase-multilingual-MiniLM-L12-v2 ⓘ	64.25	39.19	51.72	7500	420 MB
paraphrase-MiniLM-L3-v2 ⓘ	62.29	39.19	50.74	19000	61 MB

—

item1 	item2 	score 
Mountain Bike 29 Dropp 27v Freio Hidraulico	Mountain Bike 29 Dropp 27v Freio Hidraulico	1
Tv Portátil Tomate Mtm-808 Led 8 Polegadas 100v/240v	Tv Portátil Tomate Mtm-808 Led 8 Polegadas 100v/240v	1
Sapatenis Masculino Da Polo Lançamento 30% Off	Sapatenis Masculino Da Polo Lançamento 30% Off	1
Tênis Slip On	Tênis Slip On	1
Tênis Converse All Star Cano Alto	Tênis Converse All Star Cano Alto	1
Sapatênis Tênis Masculino Couro Legítimo Cano Alto Confort	Sapatênis Tênis Masculino Couro Legítimo Cano Alto Confort	1
Tênis Mizuno Wave Prophecy 6 - Ref***0000039	Tênis Mizuno Wave Prophecy 6 - Ref***0000057	0.999789
Tênis Mizuno Wave Prophecy 6 - Ref***0000057	Tênis Mizuno Wave Prophecy 6 - Ref***0000054	0.999721
Tênis Mizuno Wave Prophecy 6 - Ref***0000039	Tênis Mizuno Wave Prophecy 6 - Ref***0000054	0.999683
Tênis Mizuno Wave Prophecy 6 - Refff***000552354	Tênis Mizuno Wave Prophecy 6 - Refff***0005555178	0.999544

Productos más similares en titles_test

Recomendaciones



Validar los resultados de similitud entre productos.

Limpieza previa en los nombres



Agregar a los títulos las **imágenes** de producto y usar modelos como **clip-ViT-B-32-multilingual-v1**



Definir un **límite** para saber si los productos son iguales.

Optimizar la búsqueda con **Árboles KD**

Progreso: Área temática 2

Logro 1

- Inserta tu texto aquí
Inserta tu texto aquí
Inserta tu texto aquí
- Inserta tu texto aquí
Inserta tu texto aquí
Inserta tu texto aquí.

Logro 2

- Inserta tu texto aquí
Inserta tu texto aquí
Inserta tu texto aquí
Inserta tu texto aquí
Inserta tu texto aquí.
-

Previsión de fallas

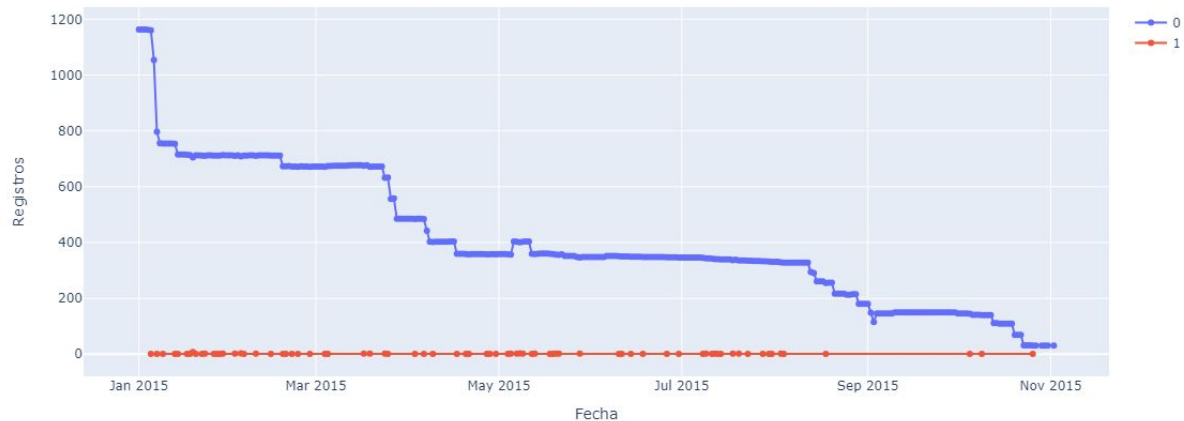


Fechas

¿El tiempo transcurrido tendrá efecto sobre la probabilidad de falla?

- Estamos obteniendo **menos información** a medida que pasa el tiempo
- **No se presentan más fallas** a medida que transcurre el año

Línea de Tiempo vs Numero de registros por Categoría de Fallo



80% Entrenamiento

20% Prueba

Conjunto de
entrenamiento

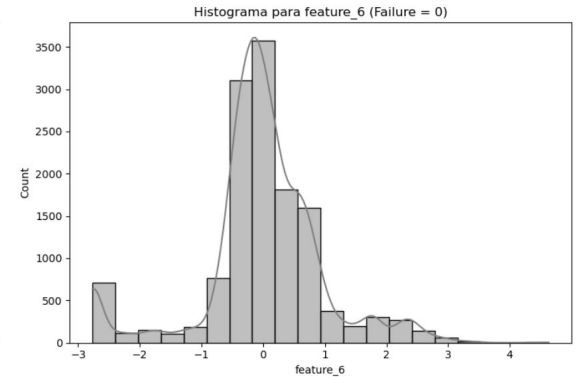
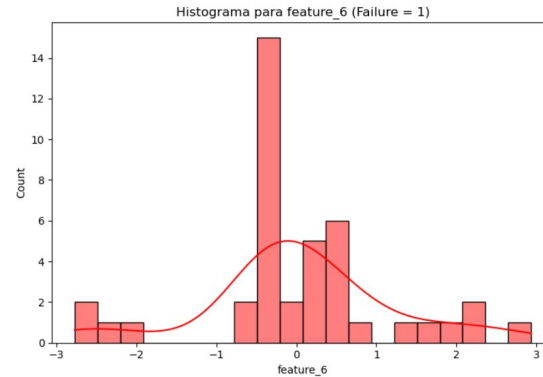
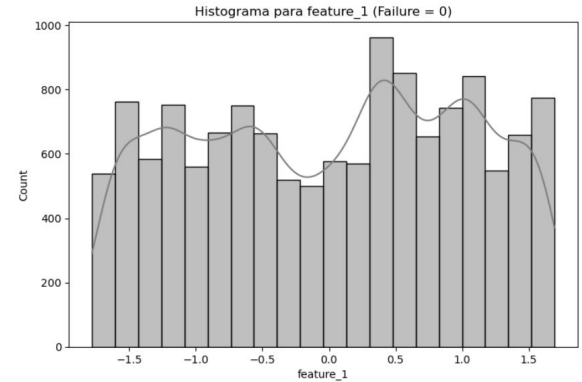
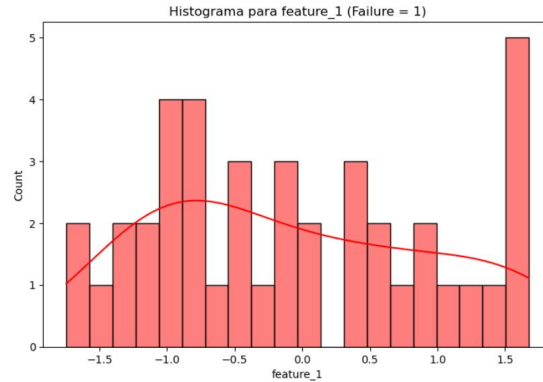


Sobremuestreo

Submuestreo

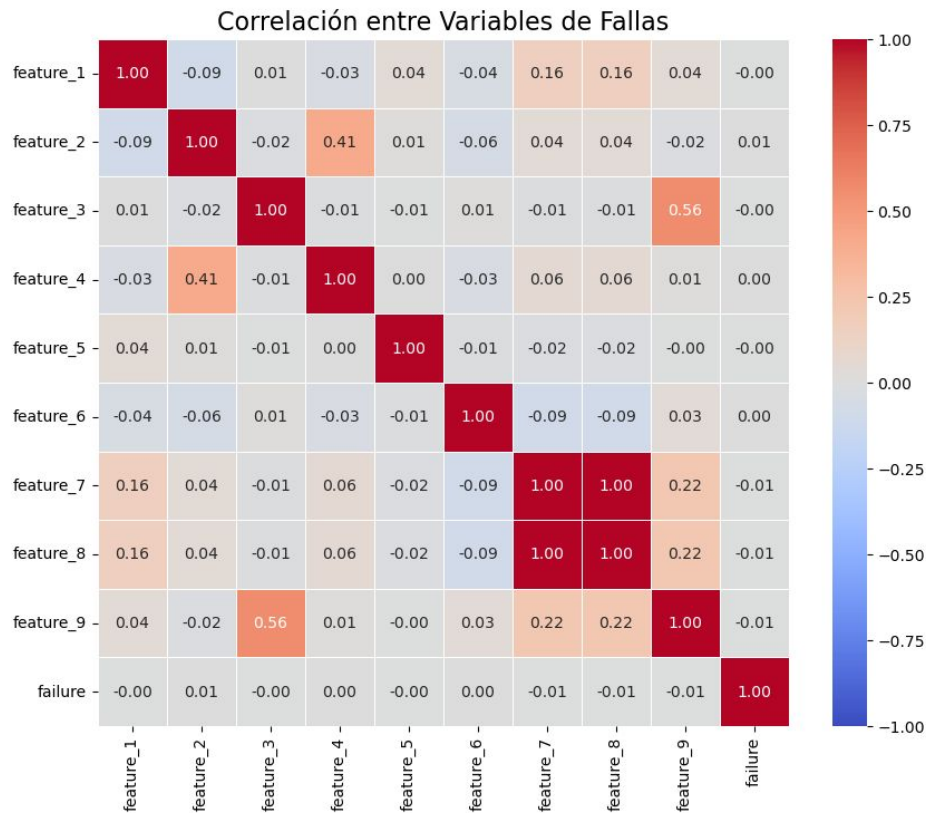
Distribuciones

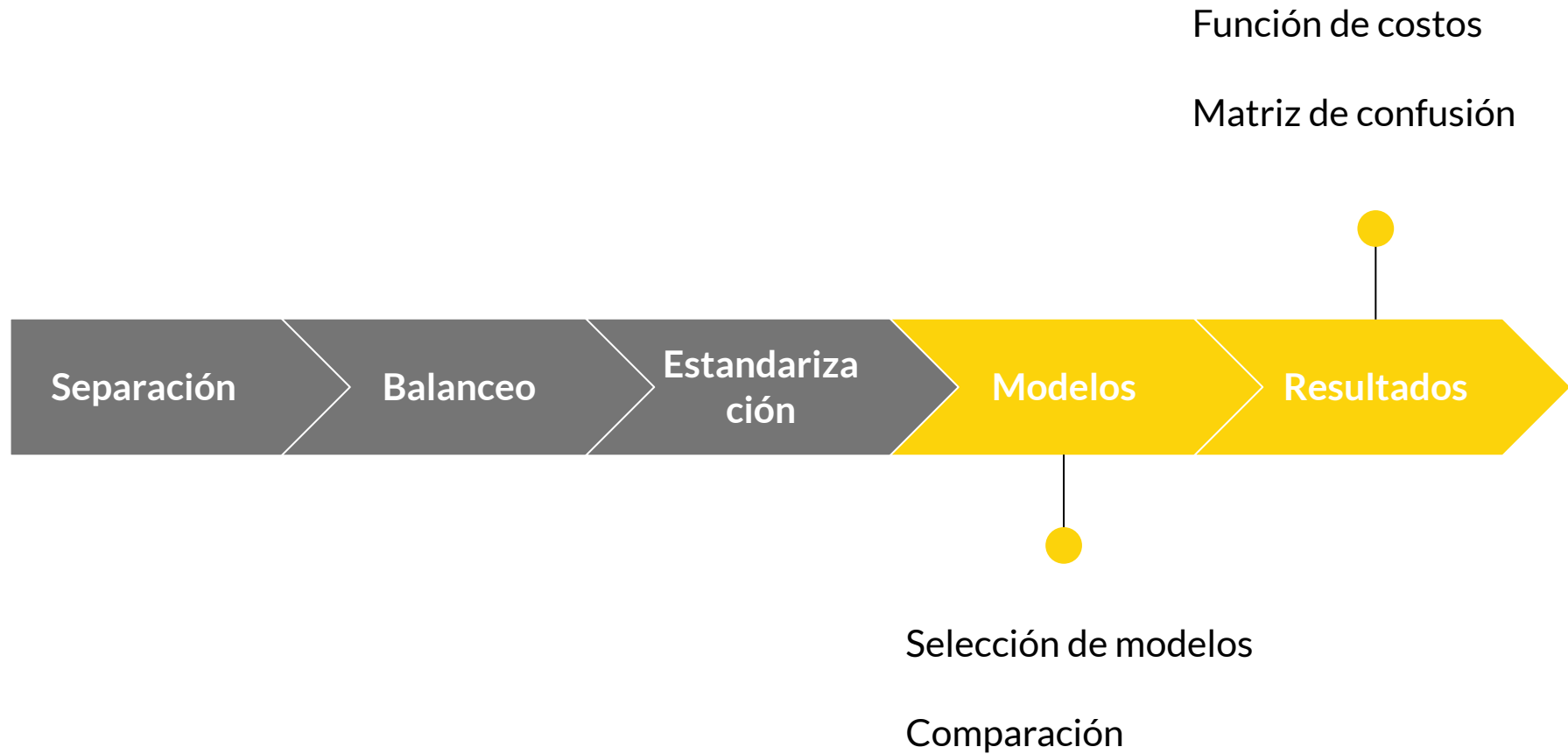
- El **desbalanceo** hace que se vean ligeras diferencias
- La **distribución** es similar en falla y no falla
- Asimetría positiva en la mayoría de atributos



Correlaciones

- Los atributos 7 y 8 aportan la misma información al modelo
- Analizando el caso en detalle, vemos que se trata de dos variables iguales
- Eliminamos el atributo 7 para entrada al modelo





Modelos de predicción



Regresión Logística

- Explicabilidad
- Clasificación binaria eficiente (recursos computacionales y convergencia)

XG Boost

- Alto rendimiento en predicciones
 - Relaciones entre características
-

Regresión Logística

- Predicción binaria asignando los pesos
- Predicción de probabilidad de falla
- Definir umbral con mínimo costo

	precision	recall	f1-score	support
0	1.00	0.97	0.98	24878
1	0.01	0.52	0.03	21
accuracy			0.97	24899
macro avg	0.51	0.75	0.51	24899
weighted avg	1.00	0.97	0.98	24899

Matriz de Confusión:

```
[[24116  762]
```

```
[   10    11]]
```

Costo Total: 391.0

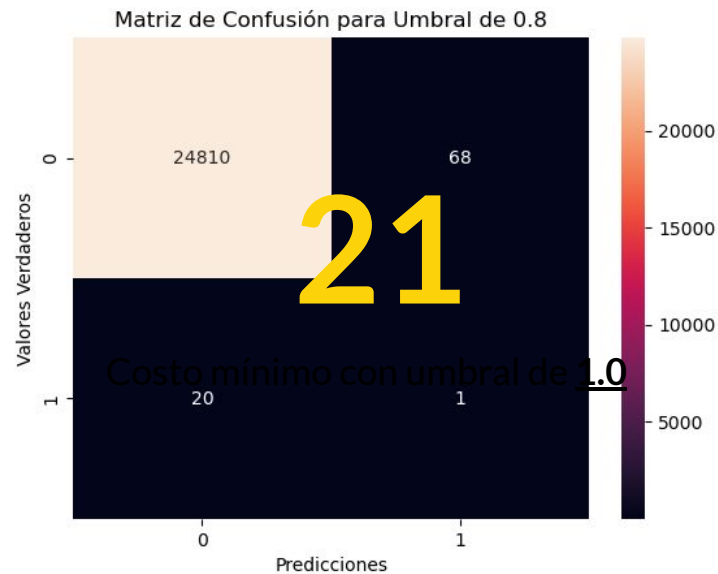
28.5

Costo mínimo con umbral de 1.0

XG Boost

- **Grid search** para optimizar la búsqueda de hiperparámetros
- Optimizar **recall**
- Buscar **umbral** según costos

```
'n_estimators': [430, 450, 500],  
'learning_rate': [0.15, 0.2, 0.25],  
'max_depth': [5, 6, 7],  
'gamma': [0, 0.05, 0.25]
```



Recomendaciones



Ajustar los
costos para las
fallas no
detectadas.



Probar modelos
de **detección de**
anomalías para
este caso



Feature
Engineering

Agregar otras
variables al
modelo usando
Target Encoding

Sugerencias

Obtener datos

- Inserta tu texto aquí

Detección de anomalías

- Inserta tu texto aquí

Revisar por tipos

- Inserta tu texto aquí
-

Pasos siguientes

Tarea 1

Inserta tu texto aquí

Inserta tu texto aquí Inserta tu texto aquí

Tarea 2

Inserta tu texto aquí

Tarea 3

Inserta tu texto aquí

iGracias!
