



MOBIBRAIN

PRESENTADO POR:

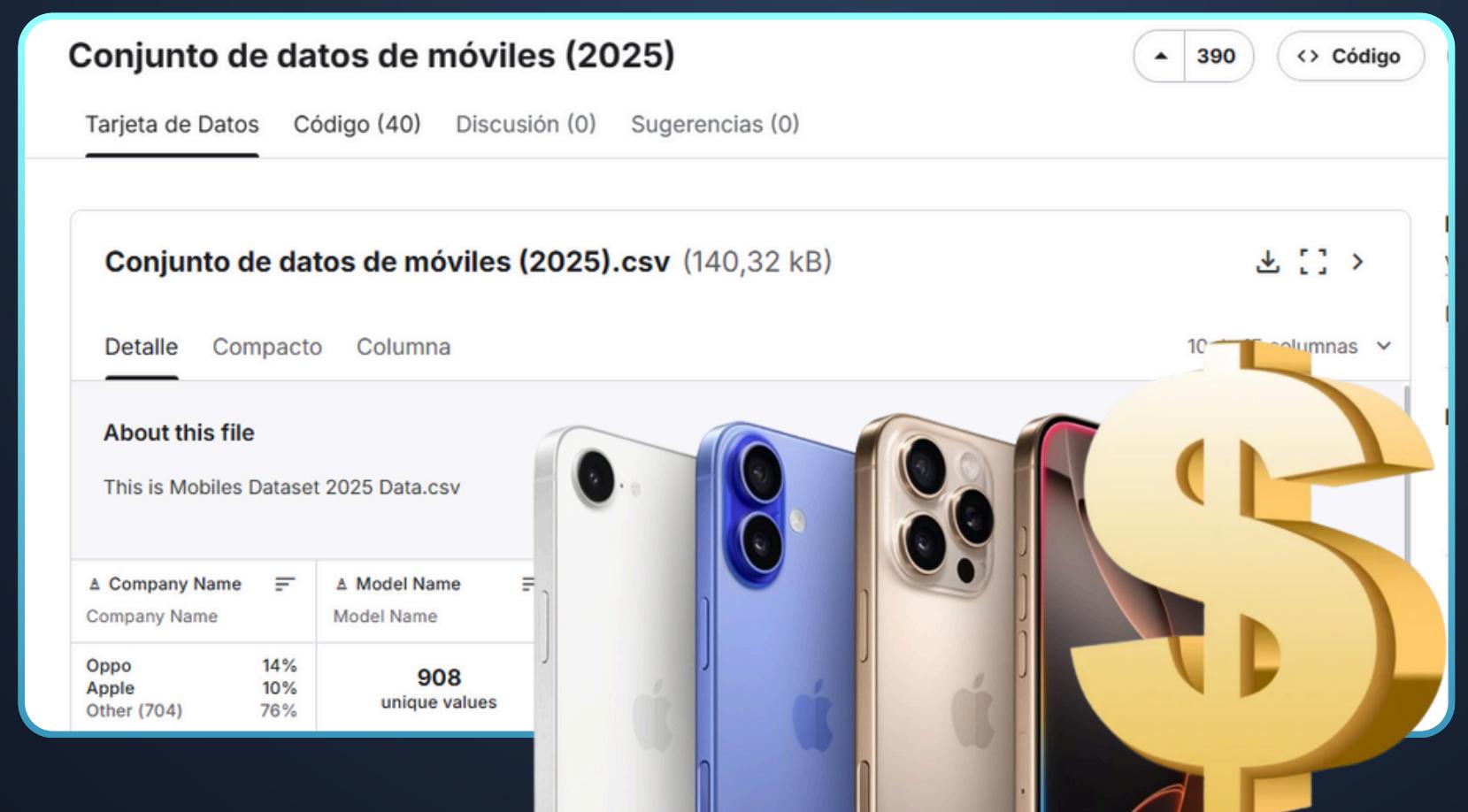
- **SOFÍA PINILLA**
- **KARINA ALFONSO**
- **KEVIN GALLARDO**



Dataset

Los dispositivos móviles generan grandes volúmenes de datos, impulsando análisis avanzados para identificar tendencias y preferencias.

En consecuencia en este proyecto se implementaron soluciones de IA para predecir tendencias del mercado (precio, batería, demanda) y identificar patrones del mercado



Dataset Preprocesamiento

Dataset utilizado

- Fuente: Mobiles Dataset 2025 (Kaggle)
- Tamaño: 930 dispositivos móviles

Desafíos iniciales del dataset

- Datos mixtos: Texto, números, valores faltantes
- Formatos inconsistentes: Especificaciones de cámara, precios con formato variable
- Variables categóricas: Procesadores de diferentes marcas

Eliminación de variables problemáticas:

- Precios de otros países (Pakistan, India, China, Dubai)
- Identificadores (Model Name, Company Name)
- Año de lanzamiento

	0
Company Name	object
Model Name	object
Mobile Weight	float64
RAM	float64
Front Camera	int64
Back Camera	int64
Processor	object
Battery Capacity	float64
Screen Size	float64
Launched Price (Pakistan)	float64
Launched Price (India)	float64
Launched Price (China)	float64
Launched Price (USA)	float64
Launched Price (Dubai)	float64
Launched Year	int64
Front Camera MP	float64
Back Camera MP	float64

dtype: object

TIPO DE PROBLEMA

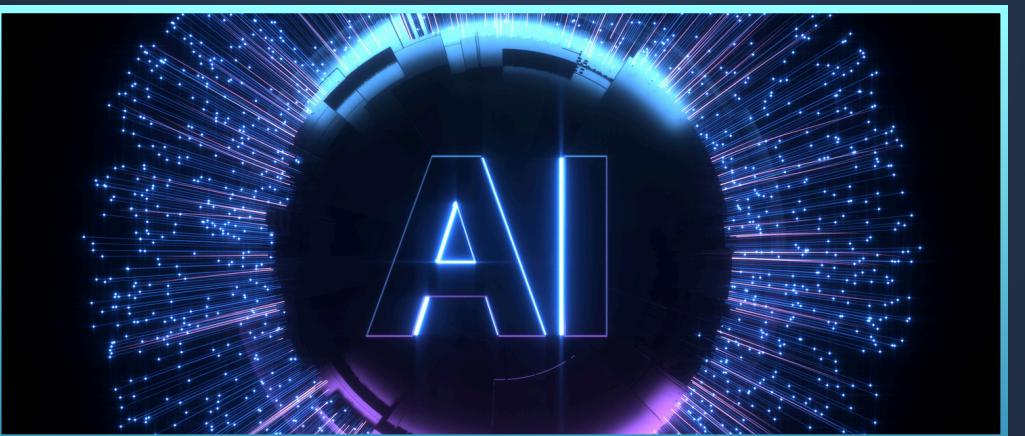
Este análisis examina diferentes modelos de regresión para predecir el precio de lanzamiento de dispositivos móviles basándose en sus características técnicas. Evaluaremos tres estimadores clásicos:

01



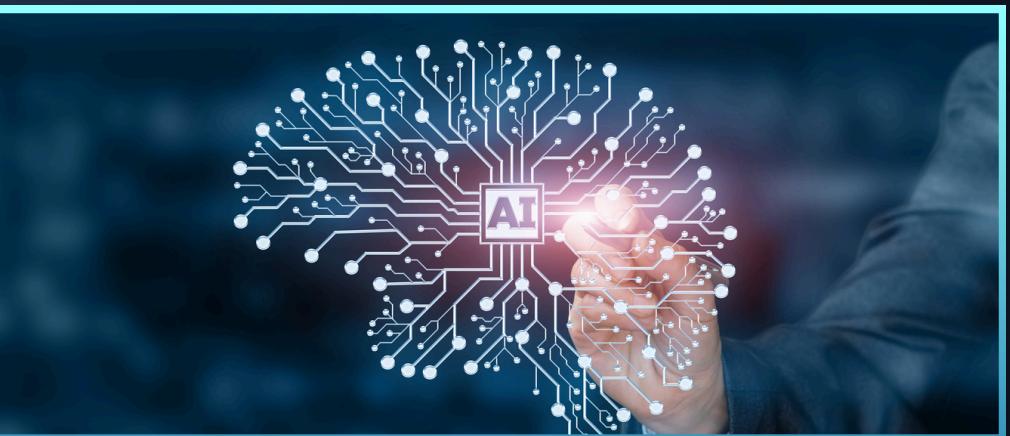
Decision Tree
Regressor

02



Random Forest
Regressor

03



Support Vector
Regressor (SVR)



Tipo de Problema y Evaluación

Básica

Nos enfrentamos a un problema de regresión, donde el objetivo es predecir un valor continuo: el precio de lanzamiento en dólares de un dispositivo móvil, basandonos en características como procesador, almacenamiento y cámara.

113.08

MAE Decision Tree

Error absoluto medio con
parámetros por defecto

92.12

MAE Random Forest

Mejor rendimiento inicial entre
los tres modelos

348.87

MAE Support Vector

Rendimiento significativamente
inferior

Decision Tree: Análisis de Profundidad

Subajuste (Underfitting)

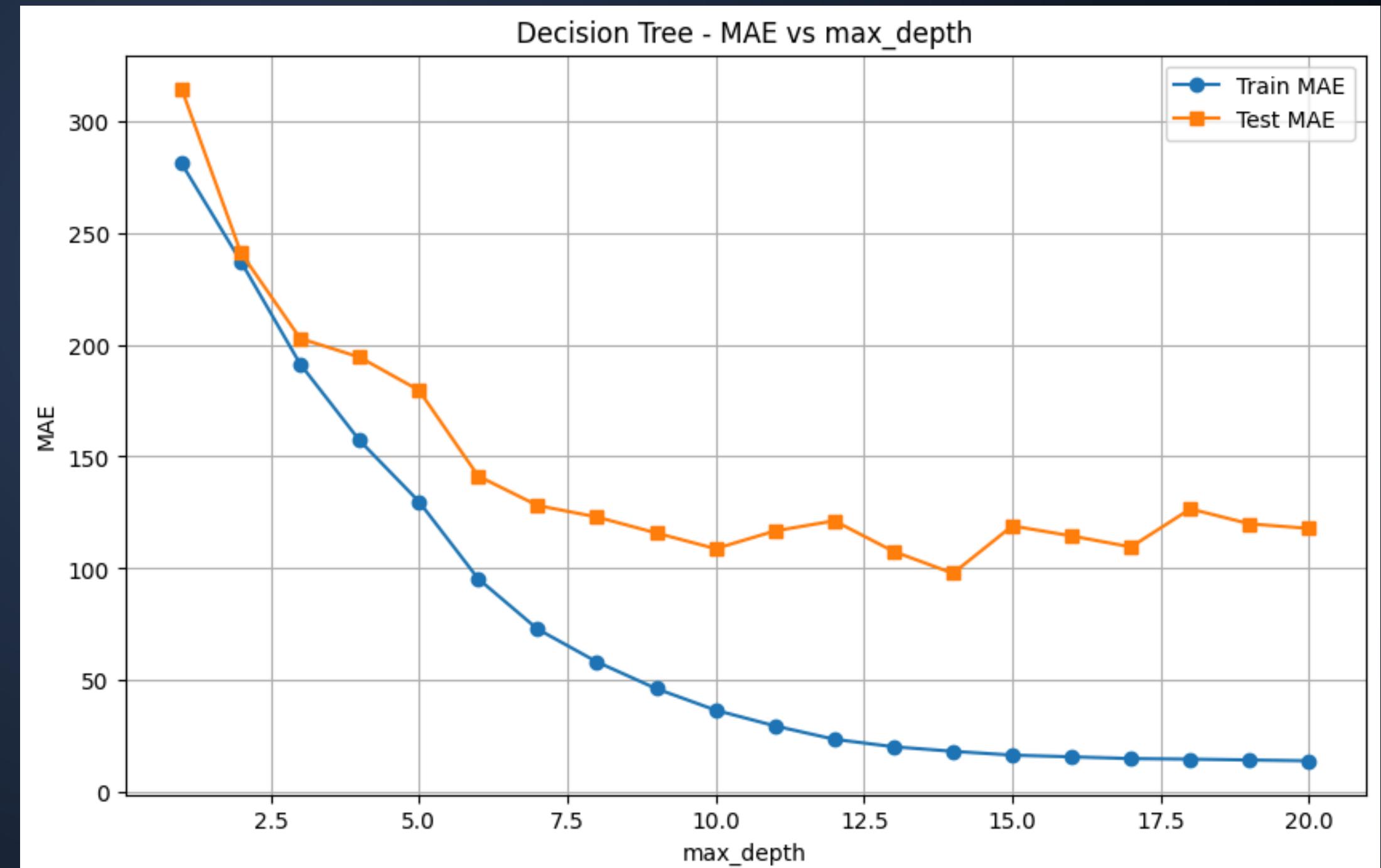
Con profundidad baja (1-5), tanto el error de entrenamiento como el de prueba son altos

Óptimo (Optimum)

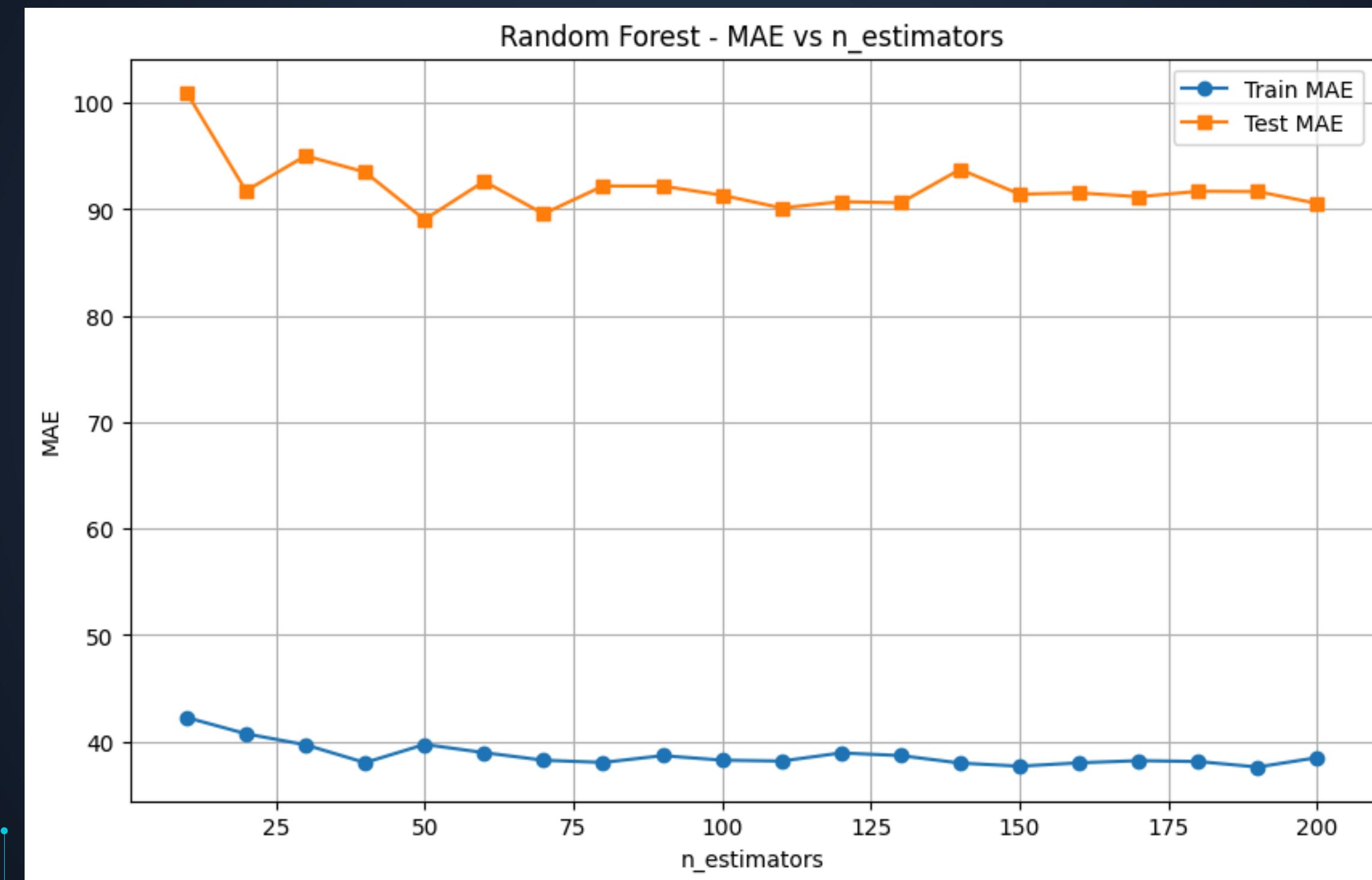
En rango intermedio (7-12), el error de prueba es mínimo con diferencia manejable respecto al entrenamiento

Sobreajuste (Overfitting)

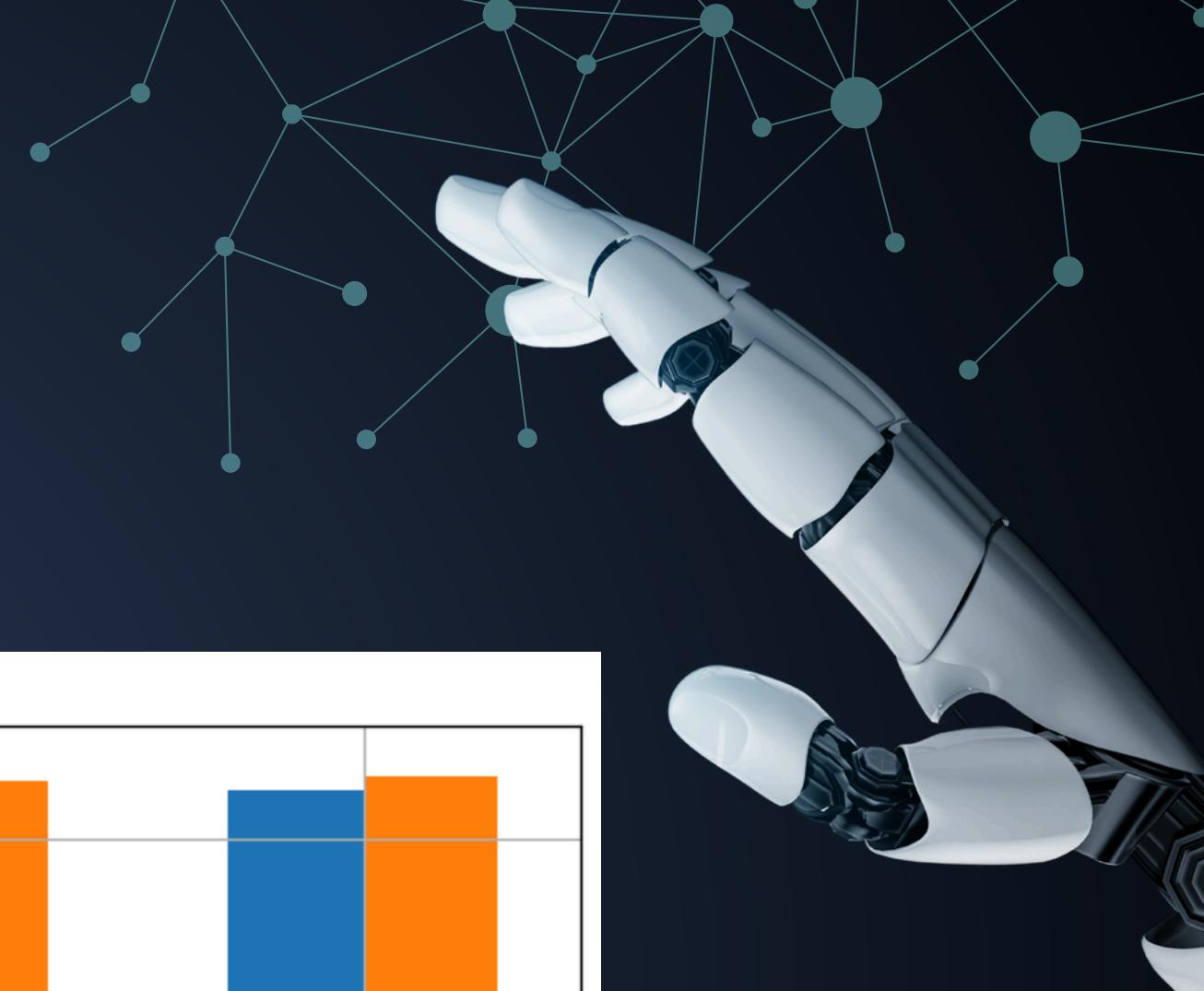
Con profundidad alta (>12), el error de entrenamiento es muy bajo pero el de prueba se mantiene alto o empeora



Random Forest: Número de Estimadores

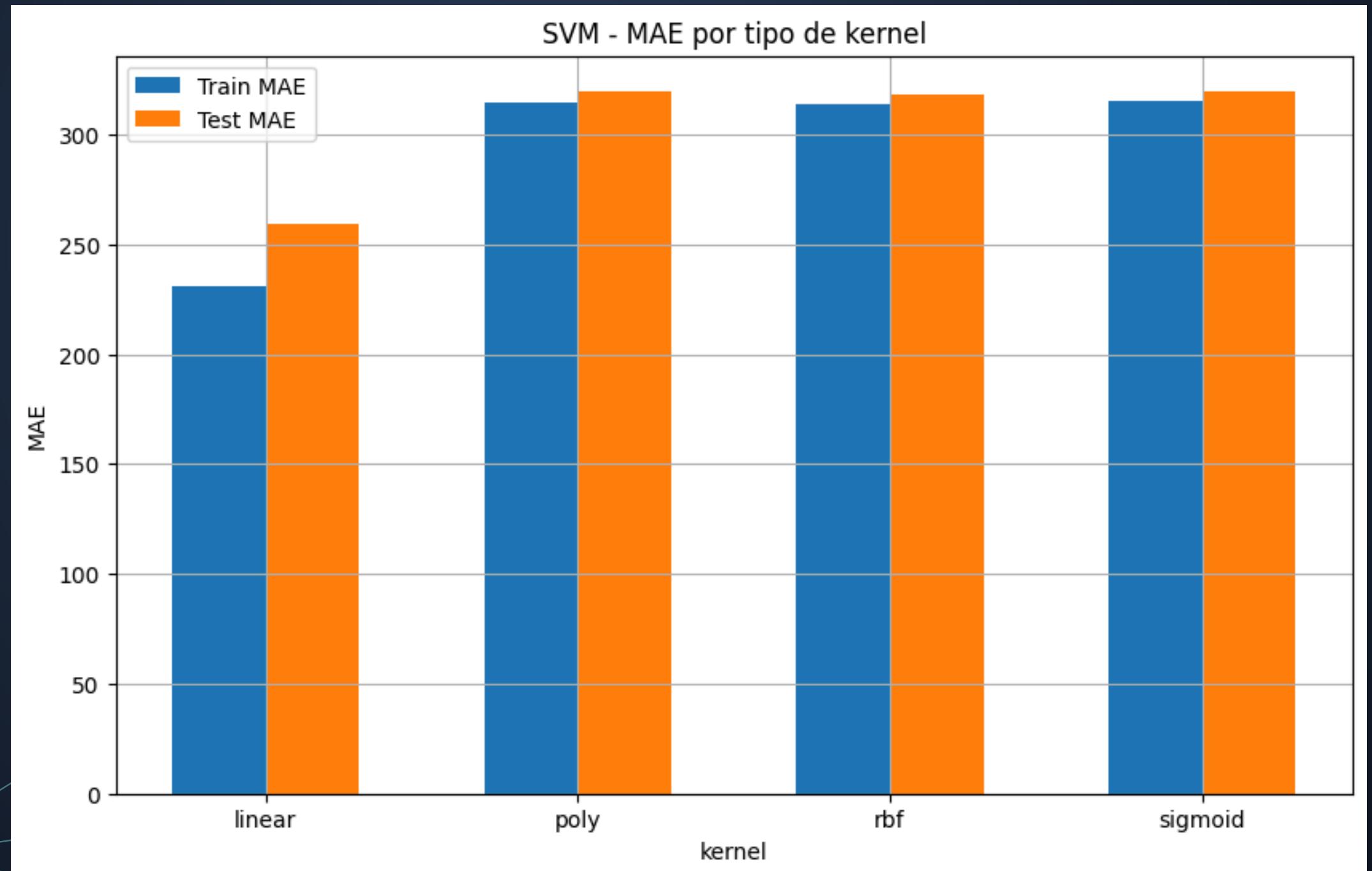


Support Vector Machine: Comparación de Kernels



Configuración Óptima

Basado en la comparación de MAE por tipo de kernel, el kernel linear parece ser la opción más prometedora para este modelo SVM en la predicción de precios de móviles.



Validación Cruzada para Decision Tree

Reducción Lineal

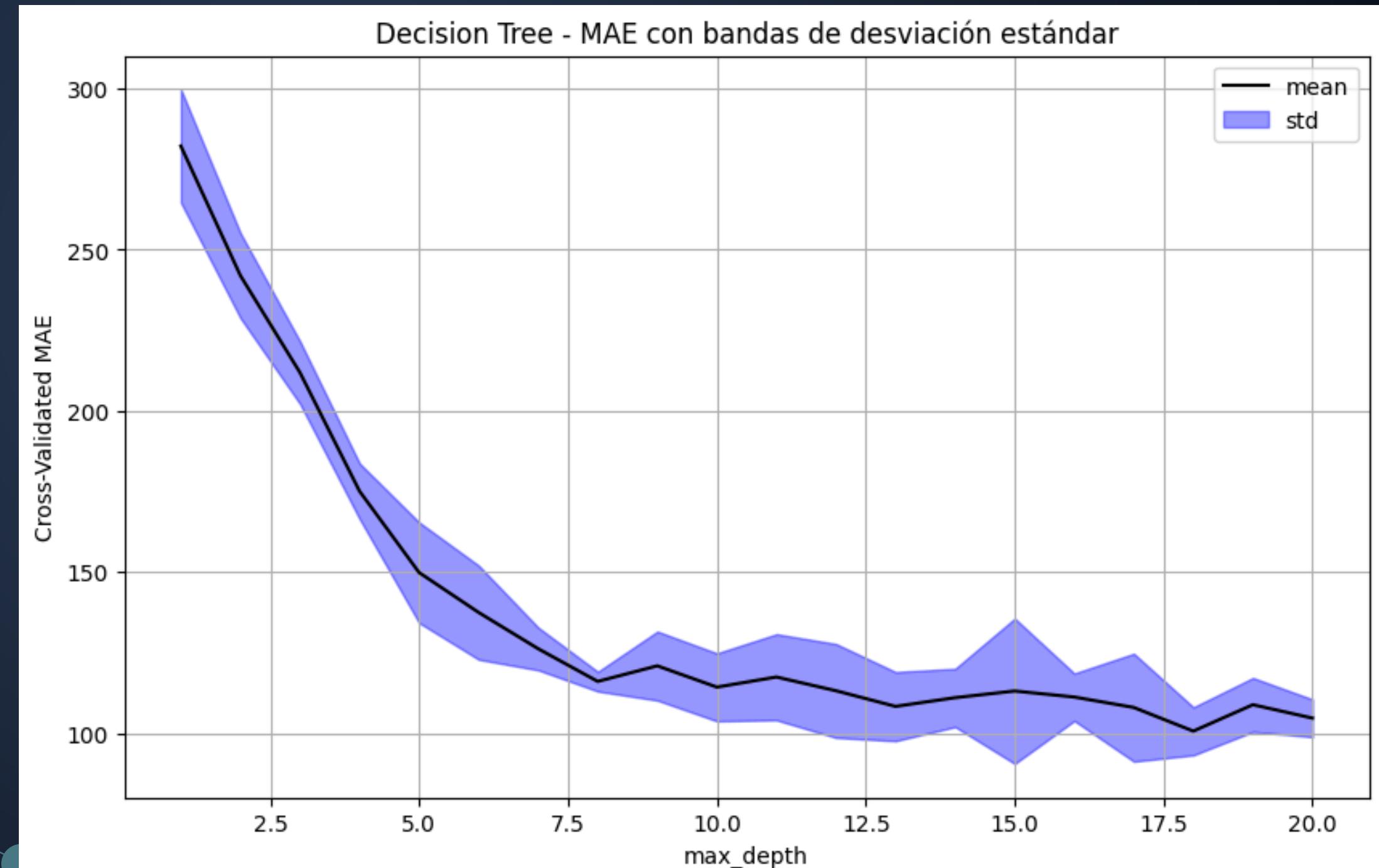
El error promedio baja significativamente al aumentar la profundidad

Estabilización

Después de profundidad 8-10, el error se estabiliza o sube ligeramente

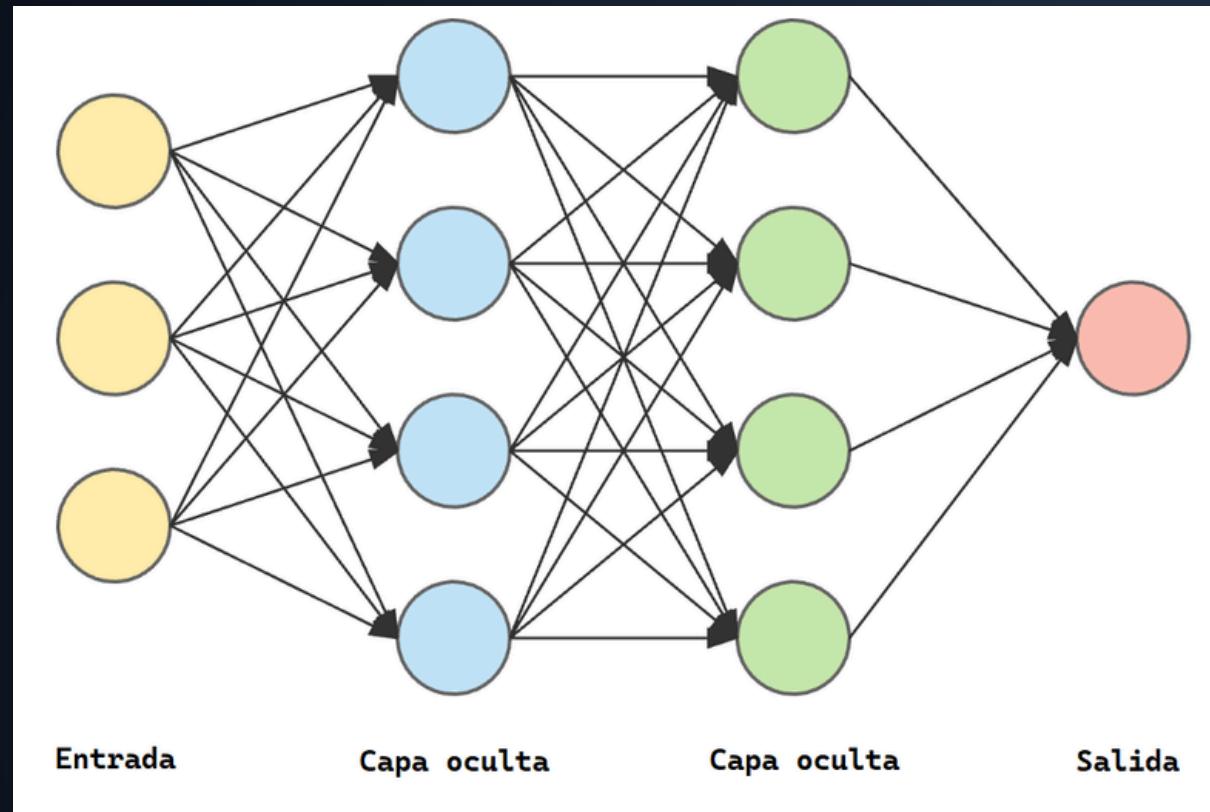
Rango Óptimo

Entre 8-14 se encuentra el mejor balance entre error bajo y variación reducida



Redes Neuronales

Configuraciones Aplicadas



Se implementaron tres arquitecturas de redes neuronales densas

- **Red Neuronal 1:** 3 capas ocultas, 128 neuronas por capa, activación ReLU
- **Red Neuronal 2:** 6 capas ocultas, 128 neuronas por capa, activación ReLU
- **Red Neuronal 3:** 10 capas ocultas, 128 neuronas por capa, activación ReLU

Todas las redes fueron entrenadas con:

- **10 epochs** utilizando el **optimizador Adam** con **learning rate de 0.001**
- Función de pérdida **MSE** (Mean Squared Error) y **MAE** (Mean Absolute Error)

Nota: Se aplicó normalización StandardScaler a X e Y. Ademas se incluyo Dropout del 0.2 para prevenir overfitting

Redes Neuronales

Resultados Obtenidos

RESUMEN COMPARATIVO DE MODELOS			
Modelo	MSE	MAE	RMSE
3_capas	65771.61	180.02	256.46
6_capas	71422.10	191.52	267.25
10_capas	106202.22	265.49	325.89

Mejor modelo (menor MSE): 3_capas
MSE: 65771.61
MAE: 180.02
RMSE: 256.46

Ejemplos de predicciones vs valores reales (mejor modelo):		
Real	Predicho	Error
379.00	954.58	575.58
139.00	195.45	56.45
1099.00	1072.63	26.37
549.00	619.66	70.66
749.00	801.82	52.82
699.00	620.14	78.86
799.00	1127.30	328.30
230.00	359.34	129.34
360.00	180.79	179.21
199.00	266.98	67.98

Consideraciones:

- Modelo de 3 capas tuvo el mejor rendimiento con un MAE de \$180.02.
- Modelos más complejos mostraron peor desempeño.
- MAE de \$180: Predicciones se desvían, en promedio, \$180 del precio real.
- Precios del dataset: entre \$139 y >\$1,099; error relativo de 20-36%, aceptable para modelo inicial.

PCA y Clustering - K-Means y DBScan

Reducción de dimensionalidad

```
Forma original del dataset: (930, 224)
```

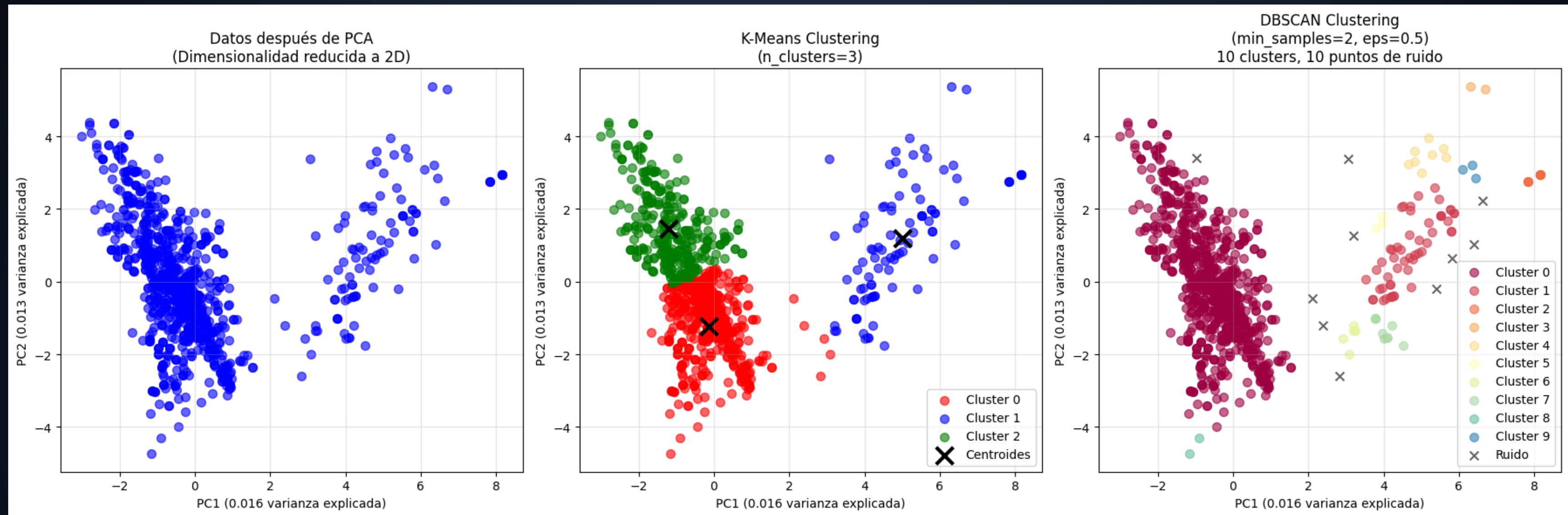
```
Forma después de PCA: (930, 2)
```

```
Varianza explicada por componente: [0.01620855 0.01266022]
```

```
Varianza total explicada: 0.0289
```

- **PCA aplicado:** Redujo 224 características a 2 componentes principales para visualizar clustering.
- **Varianza explicada:** PC1 (1.6%) y PC2 (1.3%), total 2.9%, baja debido a muchas variables categóricas (one-hot encoding).
- **Interpretación:** Datos muy distribuidos e independientes, típico en datasets con muchas categorías.

PCA y Clustering - K-Means y DBSCAN



K-Means (3 clústeres):

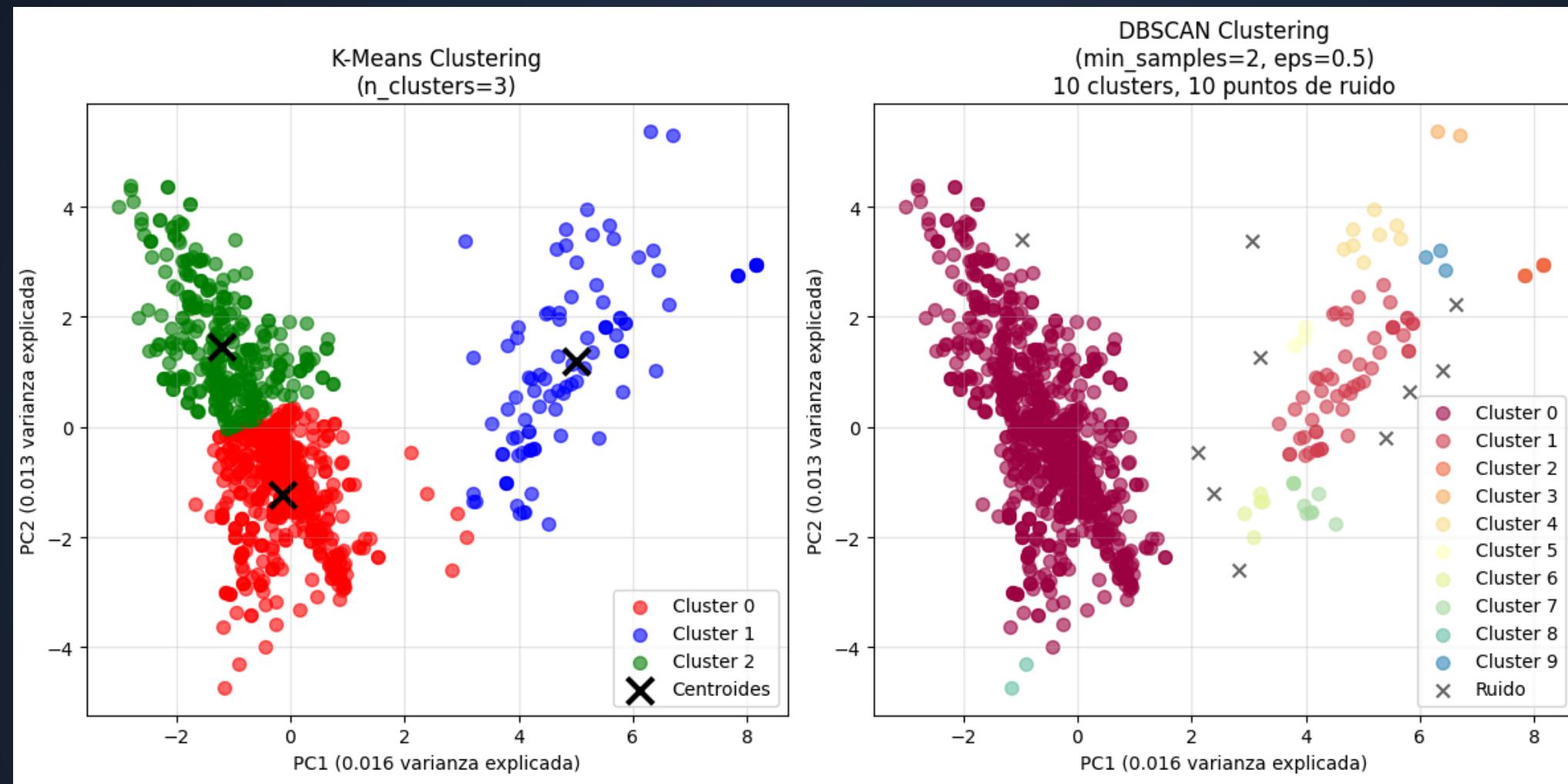
- Segmentó el mercado de móviles en 3 grupos predefinidos.
- Resultados:
 - Cluster 0: 499 dispositivos (53.7%).
 - Cluster 1: 94 dispositivos (10.1%).
 - Cluster 2: 337 dispositivos (36.2%).
- Distribución: Relativamente balanceada, con un clúster dominante (53.7%).

DBSCAN (min_samples=2, eps=0.5):

- Detectó estructura natural sin fijar número de clústeres.
- Resultados:
 - 9 clústeres naturales.
 - Clúster dominante: 828 dispositivos (89.0%).
 - 8 microclústeres: 2-52 dispositivos cada uno.
 - Ruido: 12 dispositivos (1.3%).

PCA y Clustering - K-Means y DBScan

Interpretación de resultados



K-Means:

- Segmentación artificial en 3 grupos, útil para análisis de mercado (gama baja, media, alta).
- Distribución: 53.7% (dominante), 36.2% (intermedio), 10.1% (menor).

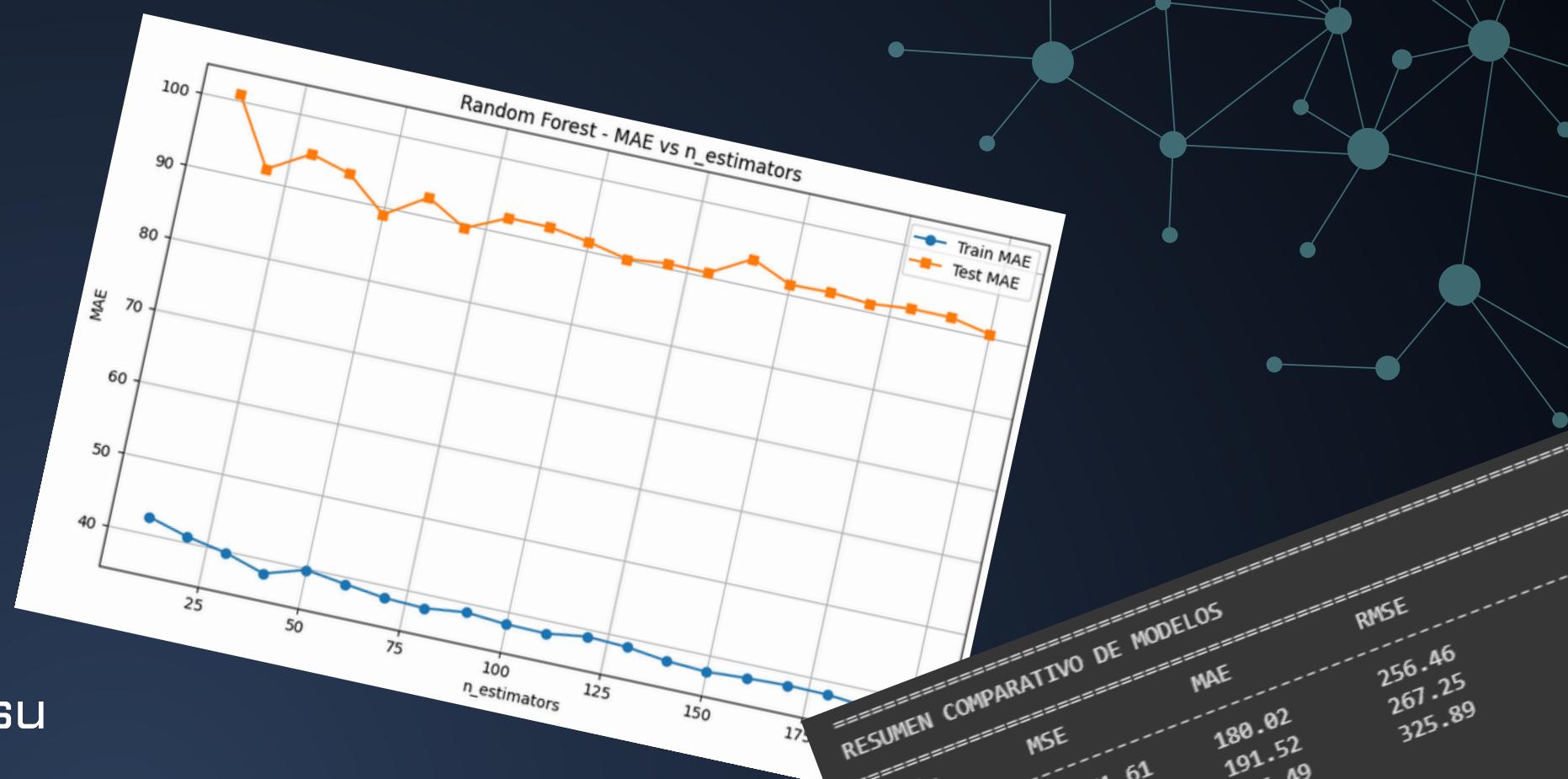
DBSCAN:

- Reveló estructura natural: 1 clúster mainstream (89%) y 8 microclústeres (nichos especializados: gama alta, gaming, etc.).
- Bajo ruido (1.3%): casi todos los dispositivos siguen patrones coherentes.

Conclusiones

Resultados Clave

- **Mejor modelo supervisado:** Random Forest, por su generalización y menor MAE.
- **Redes neuronales:** Modelos simples superaron a los complejos.
- **Clustering:** DBSCAN identificó un segmento dominante y nichos, más realista que K-Means.
- **Preprocesamiento:** Limpieza y transformación de datos fueron cruciales para el desempeño.
- **Enfoque combinado:** Modelos supervisados y no supervisados ofrecen una visión integral.



RESUMEN COMPARATIVO DE MODELOS	
Modelo	MSE
3_capas	65771.61
6_capas	71422.18
10_capas	106202.22
Mejor modelo (menor MSE): 3_capas	180.02
MSE: 65771.61	191.52
MAE: 180.02	265.49
RMSE: 256.46	267.25
	325.89

