

Predicción de Temperatura en Santander, Colombia

Un proyecto que combina ciencia de datos e inteligencia artificial para anticipar la temperatura diaria en Santander.

- Leider Esteban - 2215003
- Neider Piza - 2225613
- Milton Monsalve - 2204004

Bucaramanga

21°

Nublado

Máxima: 24° Mínima: 19°

Pronóstico de lluvia como a las 11 p.m..
Ráfagas de viento de hasta 10 km/h.

Ahora	11 p.m.	12 a.m.	1 a.m.	2 a.m.
	25%	35%		
21°	20°	20°	20°	20°

PRONÓSTICO PARA 10 DÍAS

Hoy	35%	19°	24°
Mar		19°	24°

Made with GAMMA

Descripción del Dataset y Objetivos del Proyecto

Datos hidrometeorológicos

Mediciones de temperatura del aire de estaciones IDEAM en Santander.

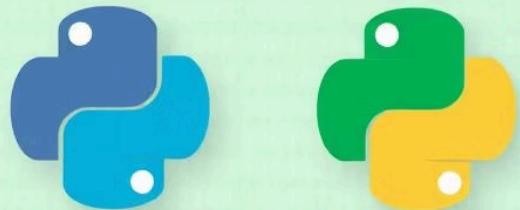
Análisis de patrones estacionales y tendencias.

Impacto crucial

- Agricultura: planificación de siembras y riegos.
- Turismo: facilita la organización de actividades.
- Salud pública: alertas tempranas de olas de calor/frío.



Preparación del Entorno y Librerías Esenciales



Google Colab

Ejecución de código
Python en la nube.



Pandas y Numpy

Manipulación y análisis
numérico de datos.



Matplotlib y Seaborn

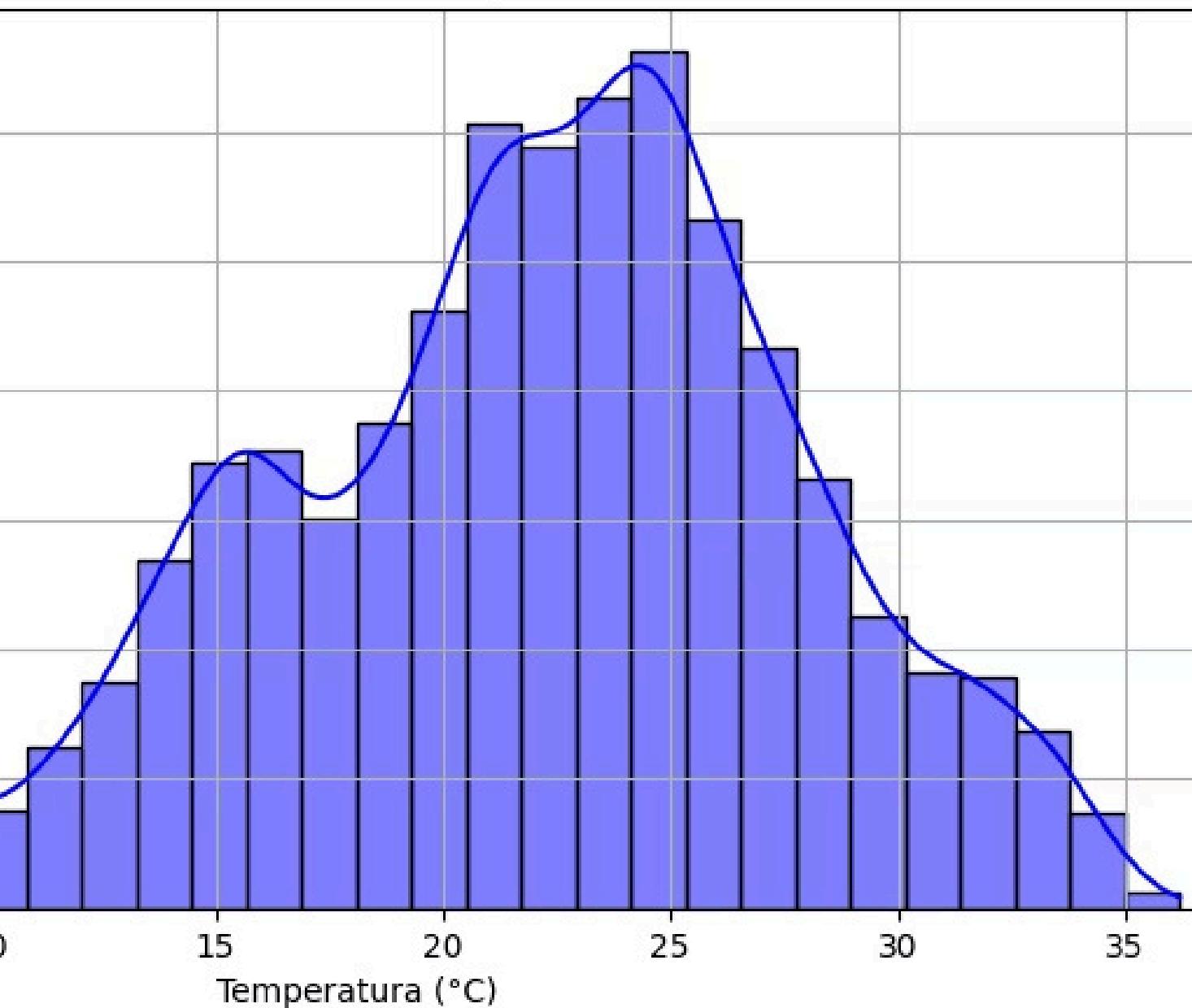
Visualización de datos
y gráficos.



Scikit-learn y TensorFlow

Modelos de ML clásico
y redes neuronales.

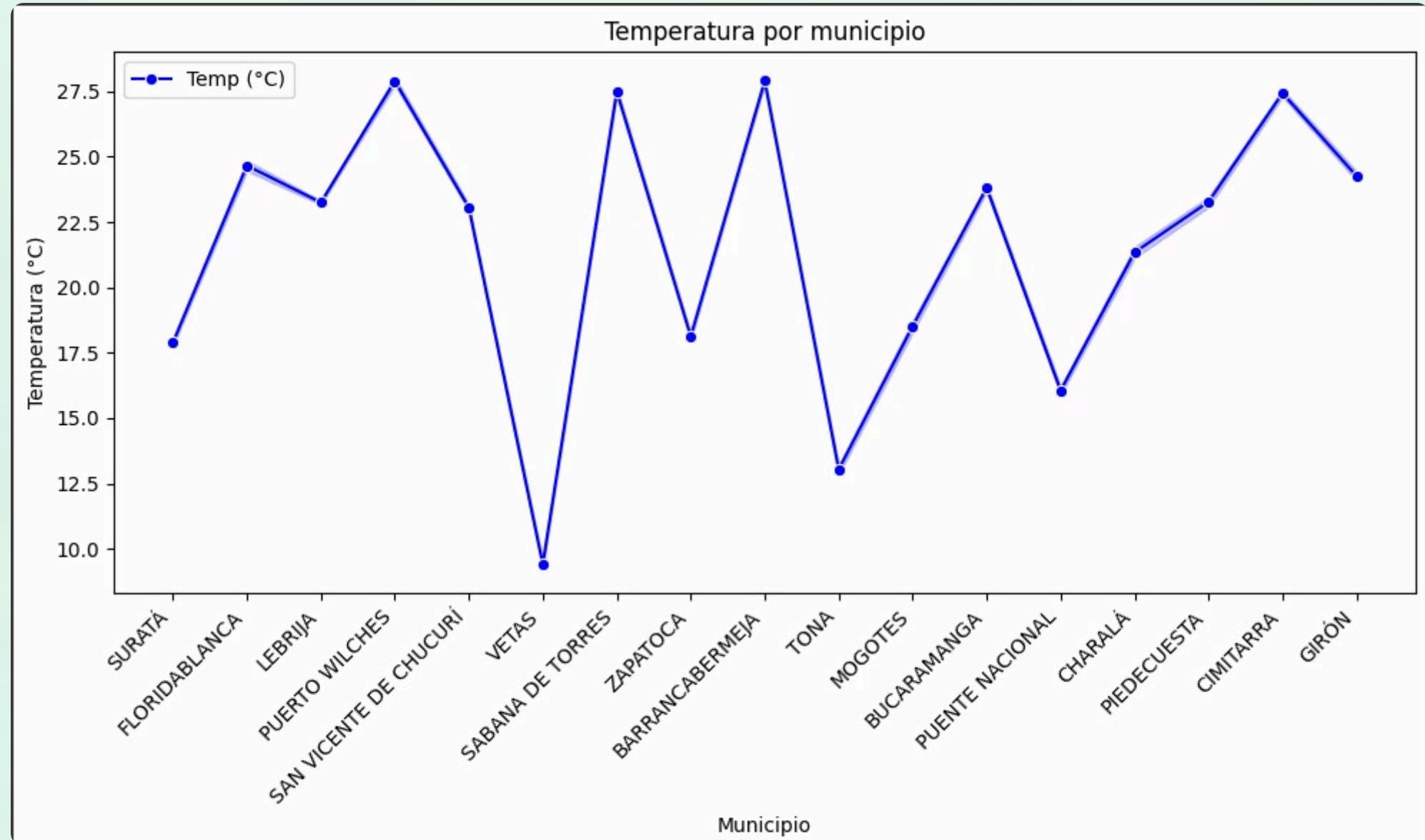
Distribución de la Temperatura

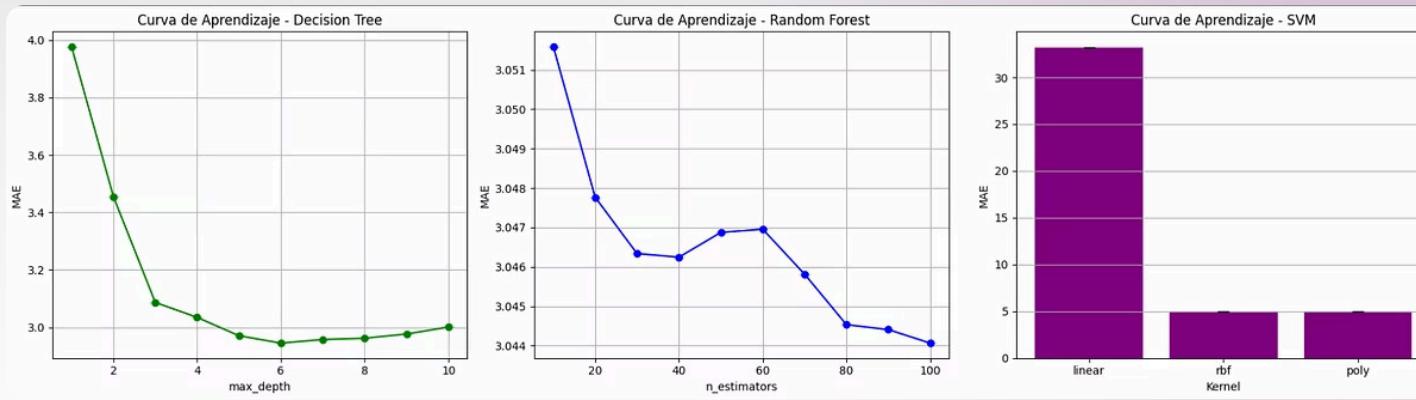


Visualización

Visualizaciones claves

- Histogramas de temperatura.
- Series temporales de temperatura.
- Temperaturas promedio por estación/municipio.
- Diagramas de dispersión espacial.





Modelado: ML Clásico



Ingeniería de características

Extracción de Día, Mes, Año, Latitud, Longitud.



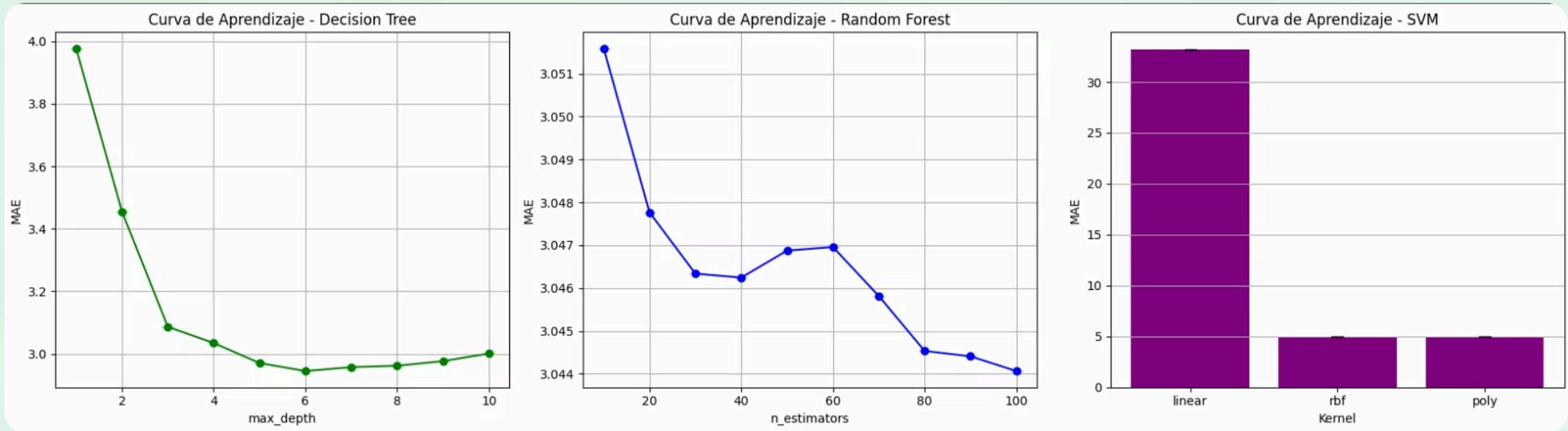
División de datos

80% entrenamiento, 20% prueba.



Modelos clásicos

Árbol de Decisión, Random Forest, SVR.



Preprocesamiento para Redes Neuronales

```
# Escalar los datos
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)

# Dividir los datos escalados
X_train_scaled, X_test_scaled, y_train_scaled, y_test_scaled = train_test_split(
    X_scaled, y, test_size=0.2, random_state=42
)
```

Modelado: Regresión y Red Neuronal

Modelo de Regresión Lineal

Establecimiento de una base predictiva.

Red Neuronal (ANN)

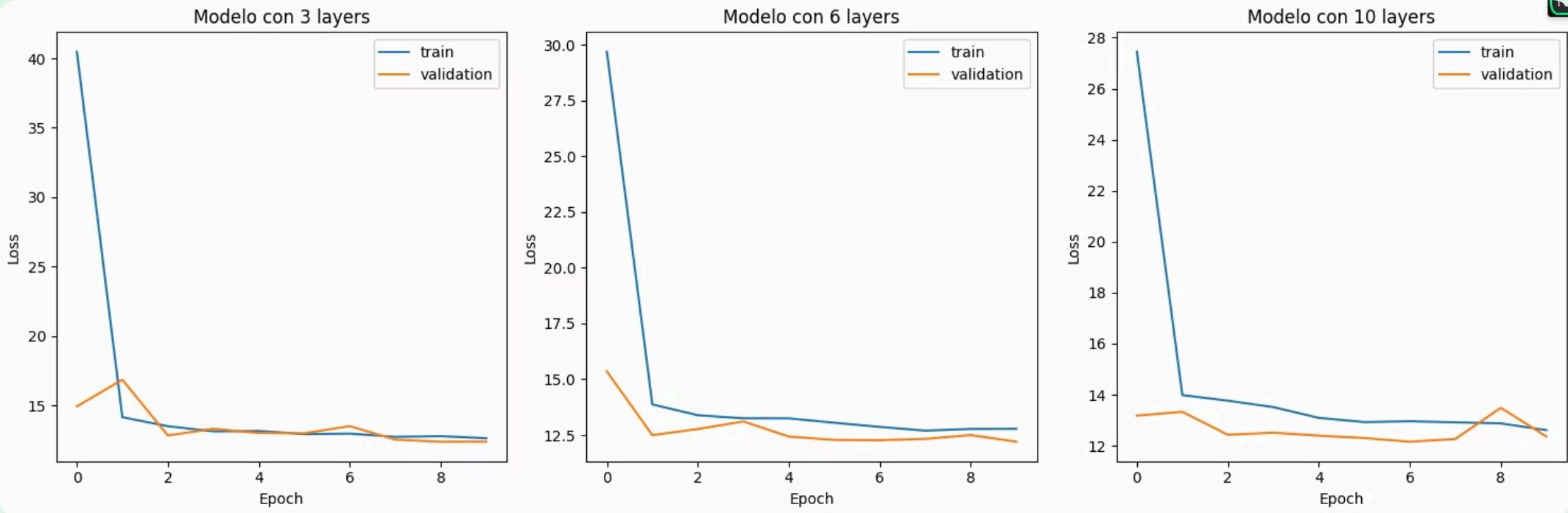
Capacidad para patrones complejos.

Métricas de Evaluación

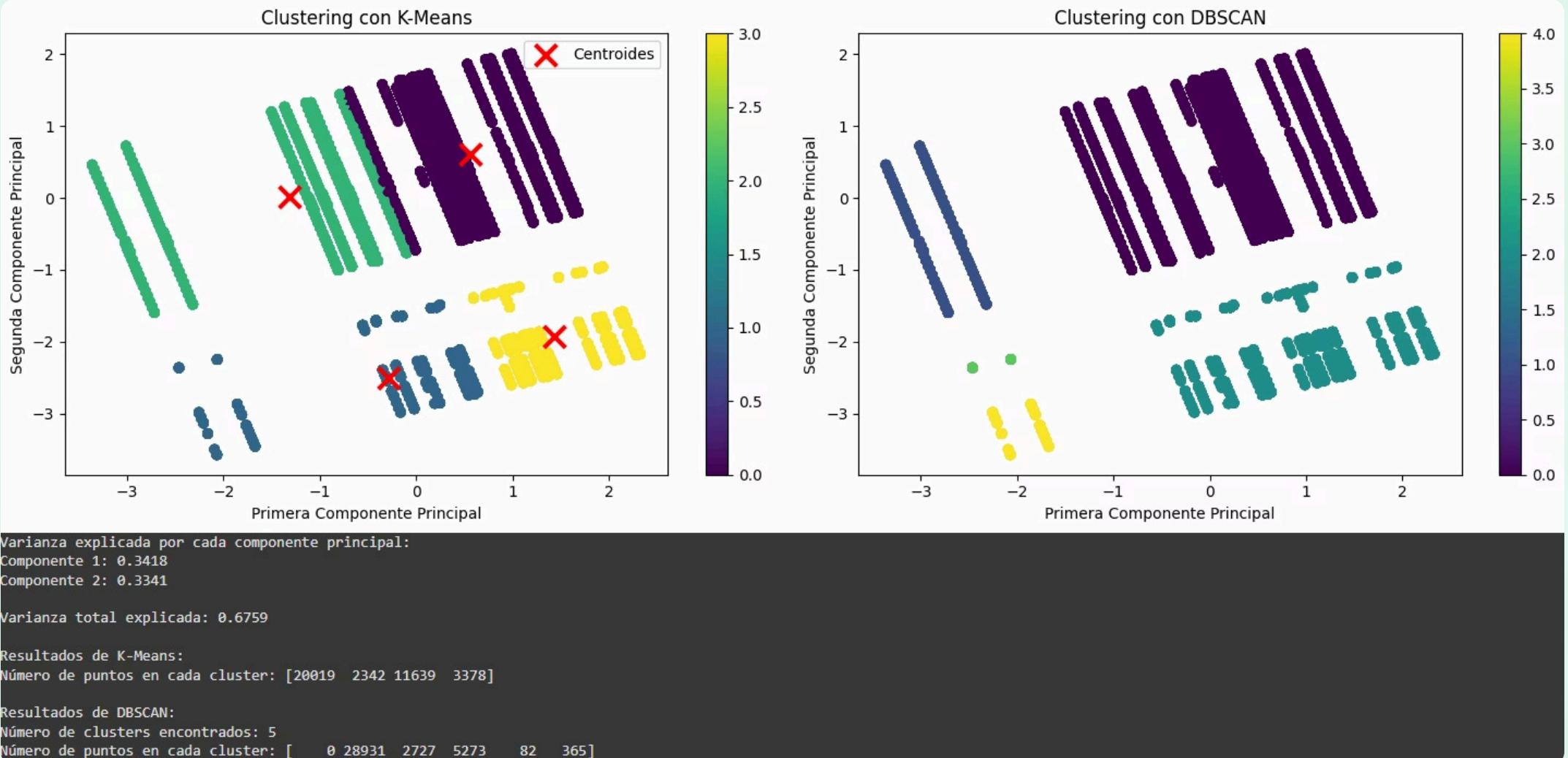
MAE, MSE, R² para precisión.

Ajuste de Hiperparámetros

Optimización para mayor exactitud.



Las curvas de aprendizaje de los modelos



Reducción de Dimensionalidad y Clustering

Resultados y Análisis Avanzado

Modelos ML Clásicos

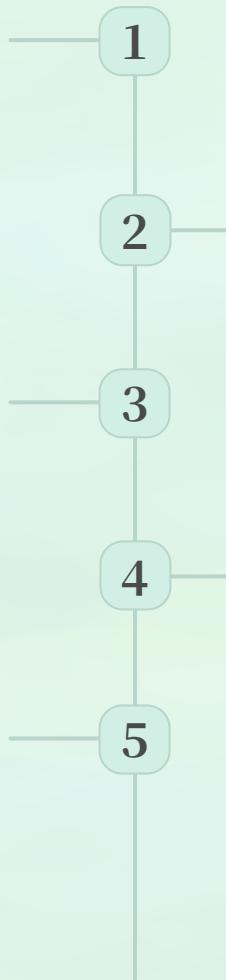
Árboles de Decisión y Random Forest alcanzaron un error medio absoluto (MAE) aproximado de 2.99°C, demostrando buen ajuste en los datos de temperatura de Santander.

Validación cruzada

Se aplicó validación cruzada de 5 folds, obteniendo una desviación estándar baja del MAE, lo que indica robustez en el rendimiento predictivo.

Análisis no supervisado

Se usó Análisis de Componentes Principales (PCA) para reducción dimensional, y técnicas de clustering como K-Means y DBSCAN para identificar agrupaciones naturales en los datos meteorológicos.



Curvas de aprendizaje

Se evaluó la variación del MAE al modificar hiperparámetros clave como **max_depth** (rango 5-20) y **n_estimators** (50-200), observando que valores óptimos mejoran la generalización del modelo.

Redes neuronales

La red neuronal profunda con 10 capas densas logró un MAE ligeramente mejor, alrededor de 2.90°C, capturando patrones no lineales de temperatura.