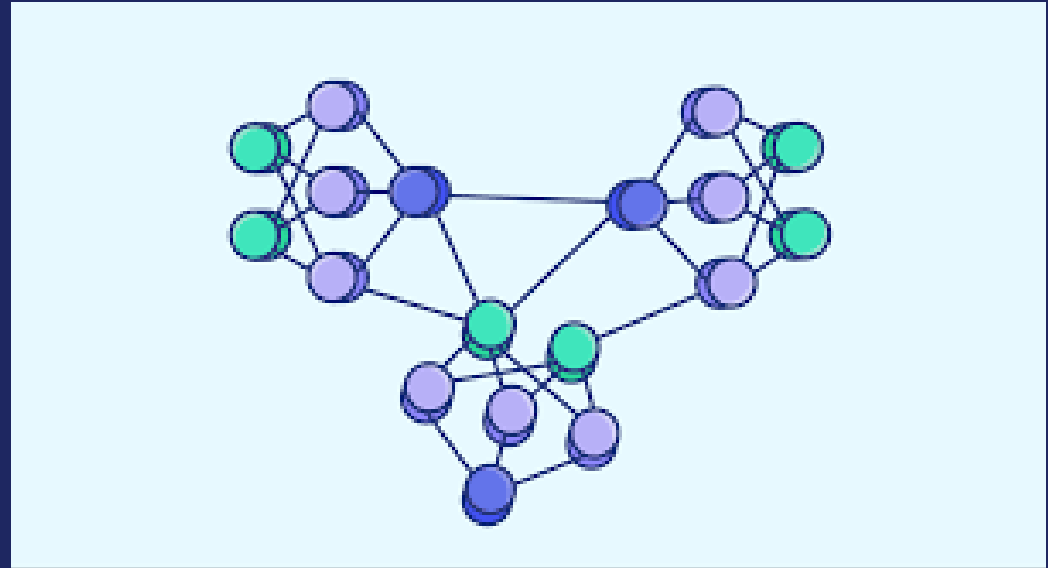


Predicción de Demanda en Retail con Redes Neuronales Recurrentes

MUBA: Deep Learning



Índice

01

Introducción y Objetivo

Motivación y pregunta de investigación

02

Dataset y EDA

Kaggle Store Item Demand Forecasting

03

Metodología

Pipeline de datos y Grid Search

04

Grid Search por modelo

SimpleRNN, LSTM, GRU, Bi-LSTM

05

Comparativa Final

Resultados con la mejor configuración de cada uno

06

Conclusiones

Hallazgos clave y limitaciones

Comparamos cuatro arquitecturas recurrentes para predecir ventas diarias en retail. Para cada modelo buscamos su mejor configuración mediante Grid Search y después comparamos lo mejor de cada uno.

SimpleRNN

Referencia base

LSTM

3 puertas, largo plazo

GRU

2 puertas, eficiente

Bi-LSTM

Bidireccional

Baseline: media móvil de 7 días para medir si los modelos DL realmente aportan algo frente a un método estadístico simple.

Hiperparámetros explorados en Grid Search:

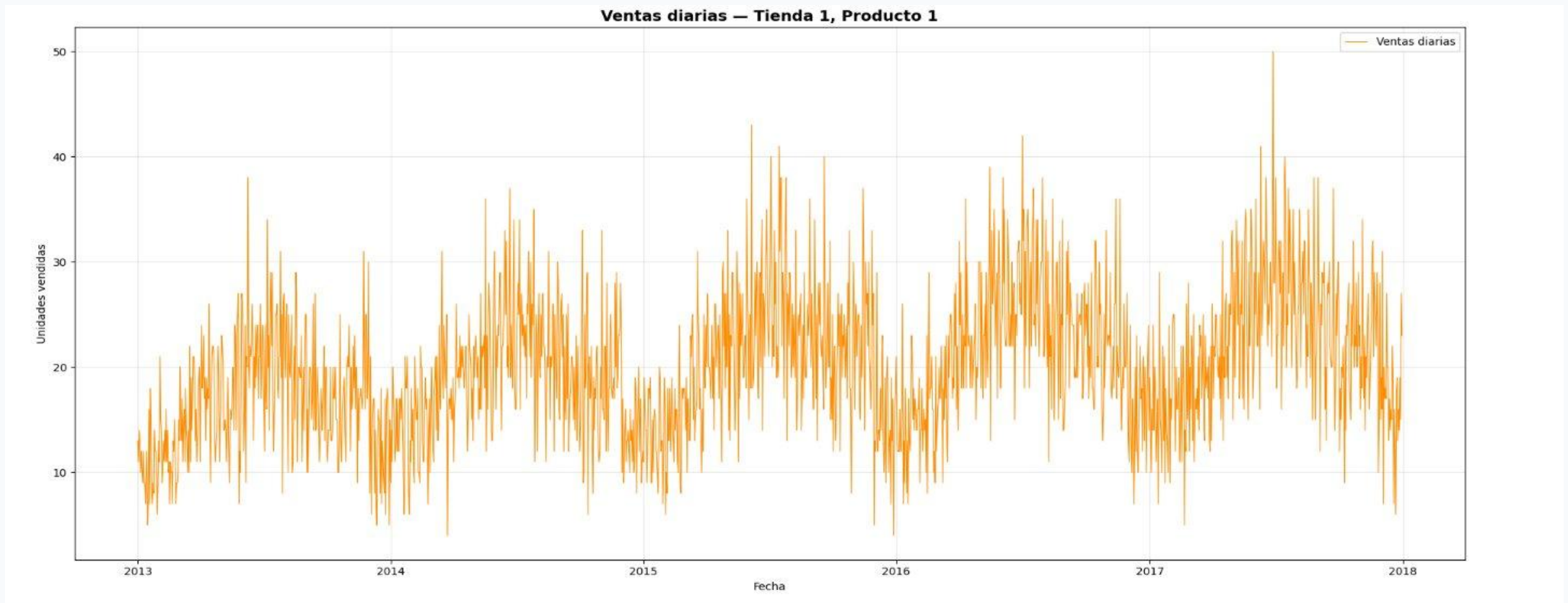
Ventana: 7, 30, 90 días · Neuronas: 32 · Activación: tanh, relu, LeakyReLU

9 combinaciones por modelo x 4 modelos = 36 entrenamientos + 16 de batch size. La rejilla completa (240) queda en el notebook.

Kaggle · Store Item Demand Forecasting Challenge

913.000 registros de ventas diarias — 10 tiendas × 50 productos × 5 años (2013–2017)

Serie de trabajo: Tienda 1 · Producto 1 (1.826 días) — Serie univariante



1

Ventana deslizante

Cogemos los últimos N días como entrada para predecir el día siguiente.

2

Split temporal 70/15/15

Train, validacion y test en orden cronologico. Los cortes van por dia, asi que los 5 modelos predicen los mismos 274 dias.

3

MinMaxScaler [0,1]

Normalizamos entre 0 y 1. El scaler se ajusta solo en train.

4

Grid Search por modelo

9 combinaciones por modelo. Se elige por MAPE en VALIDACION, nunca en test.

5

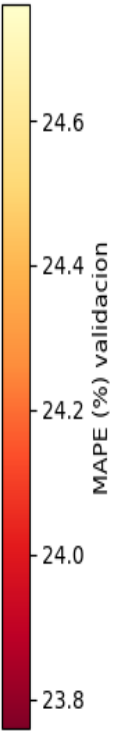
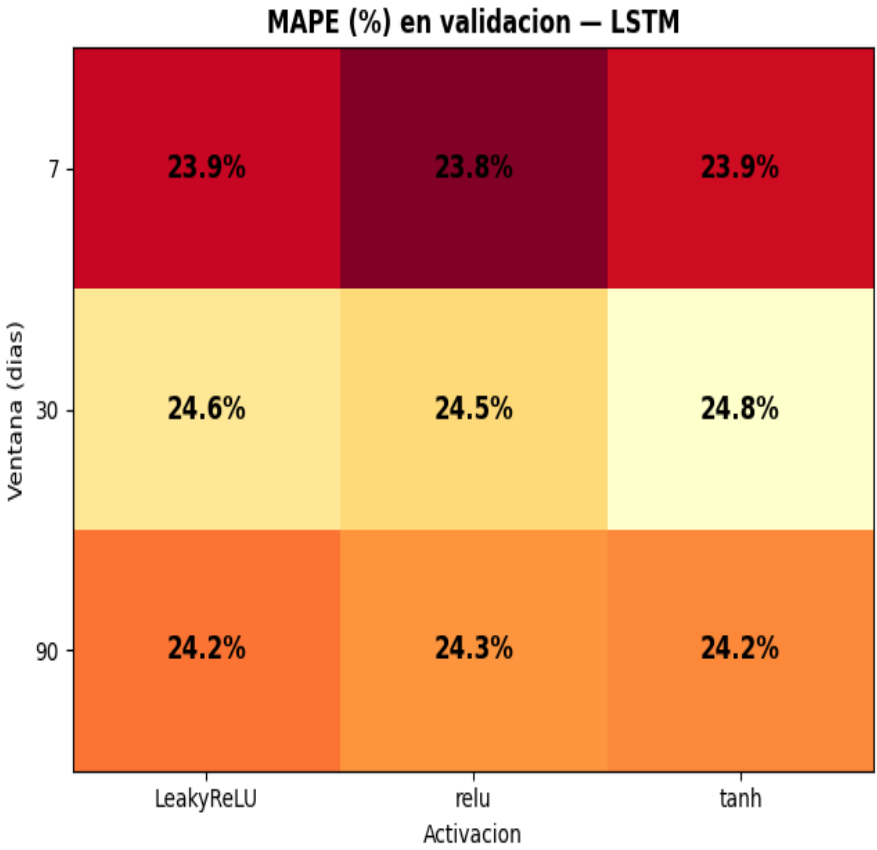
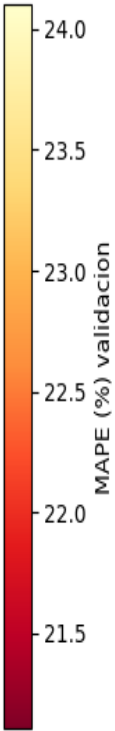
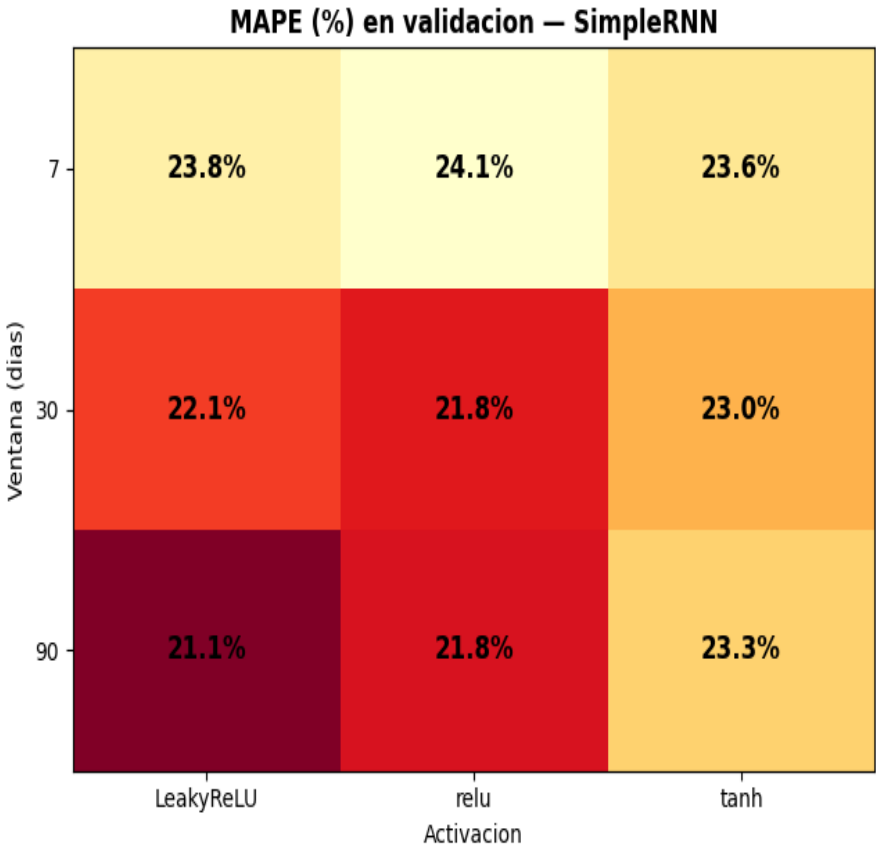
EarlyStopping

Si el error en validacion no mejora en 5 epochs, para y recupera los mejores pesos. La validacion es un tramo propio: el test no interviene.

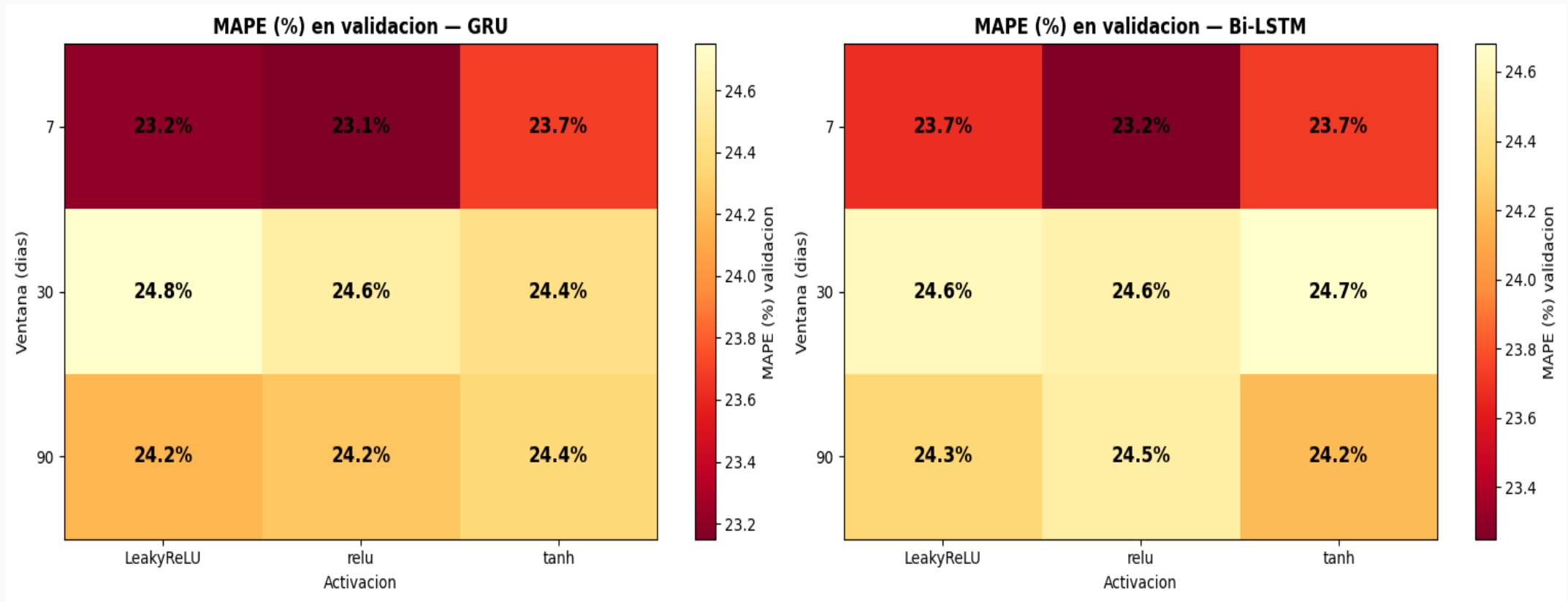
6

Comparativa final

Cada modelo con su mejor configuración + baseline de media móvil.



La SimpleRNN es la unica que prefiere ventana LARGA (90 dias, LeakyReLU, 21,11% en validacion). La LSTM prefiere ventana CORTA (7 dias, relu, 23,76%): sus puertas ya gestionan la memoria y no necesitan tanta ventana.



La GRU tambien prefiere ventana corta (7 dias, relu, 23,15%) y la Bi-LSTM igual (7 dias, relu, 23,25%). Las tres arquitecturas con puertas coinciden: el gating sustituye a la ventana larga.

Modelo	Ventana	Neuronas	Activación	MAPE validacion
SimpleRNN	90 dias	32	LeakyReLU	21,11%
LSTM	7 dias	32	relu	23,76%
GRU	7 dias	32	relu	23,15%
Bi-LSTM	7 dias	32	relu	23,25%

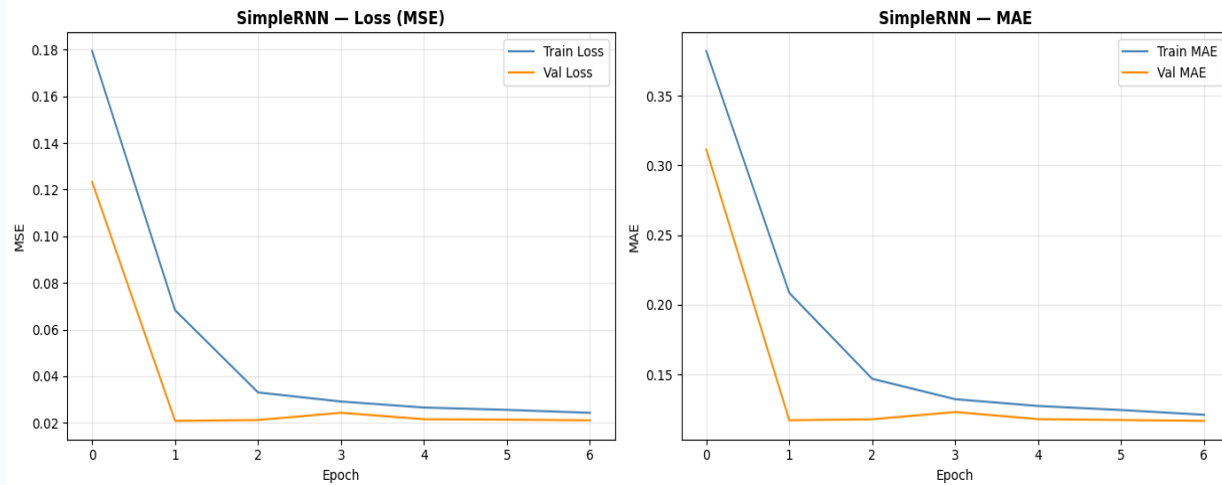
- **Hallazgos del Grid Search:**
- Las tres arquitecturas con puertas eligen ventana corta (7 dias). La SimpleRNN, que no las tiene, necesita 90: el gating sustituye a la ventana larga.
- Si hay sobreajuste: en las 36 configuraciones el MAE de validacion es peor que el de train (+0,17 a +0,52), sin una sola excepcion.
- La ventaja sobre el baseline es robusta: 49 de 52 configuraciones lo superan.

Con la mejor configuracion de cada modelo, probamos 4 batch sizes. El mejor de cada fila se elige por MAPE en validacion:

Modelo	Config	BS=16	BS=32	BS=64	BS=128	Mejor
SimpleRNN	w=90, n=32 LeakyReLU	22,01%	23,00%	21,80%	21,56%	128
LSTM	w=7, n=32 relu	23,54%	23,54%	24,55%	23,91%	16
GRU	w=7, n=32 relu	23,63%	23,21%	24,92%	24,48%	32
Bi-LSTM	w=7, n=32 relu	23,75%	23,75%	24,42%	23,64%	128

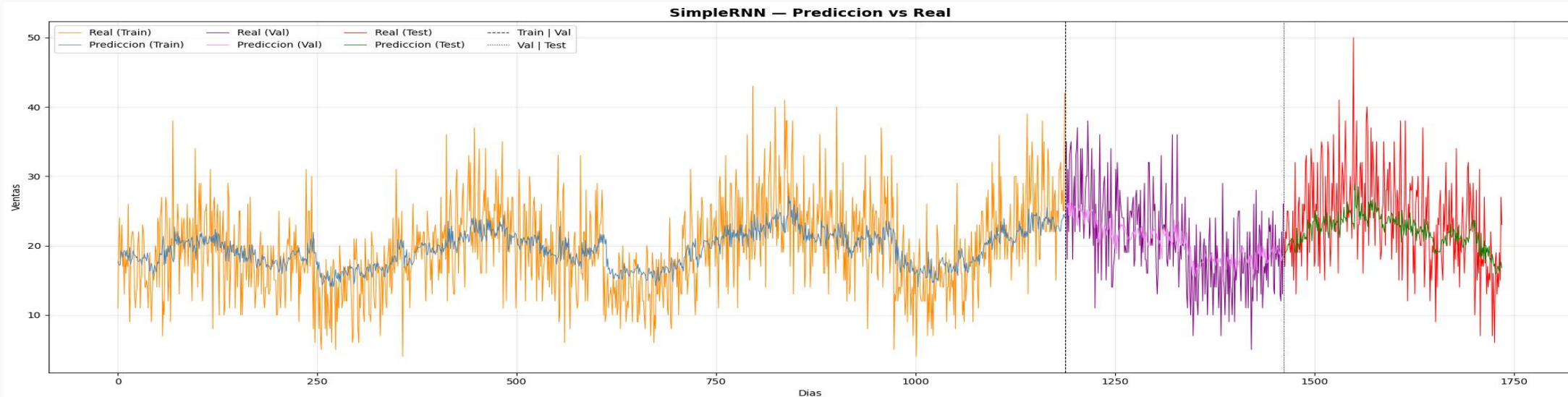
Hallazgos:
No hay una direccion unica: la LSTM prefiere 16, la GRU 32, y SimpleRNN y Bi-LSTM 128.
El impacto del batch size es menor que el de la ventana o la activacion.
Tambien se decide con validacion, no con test.

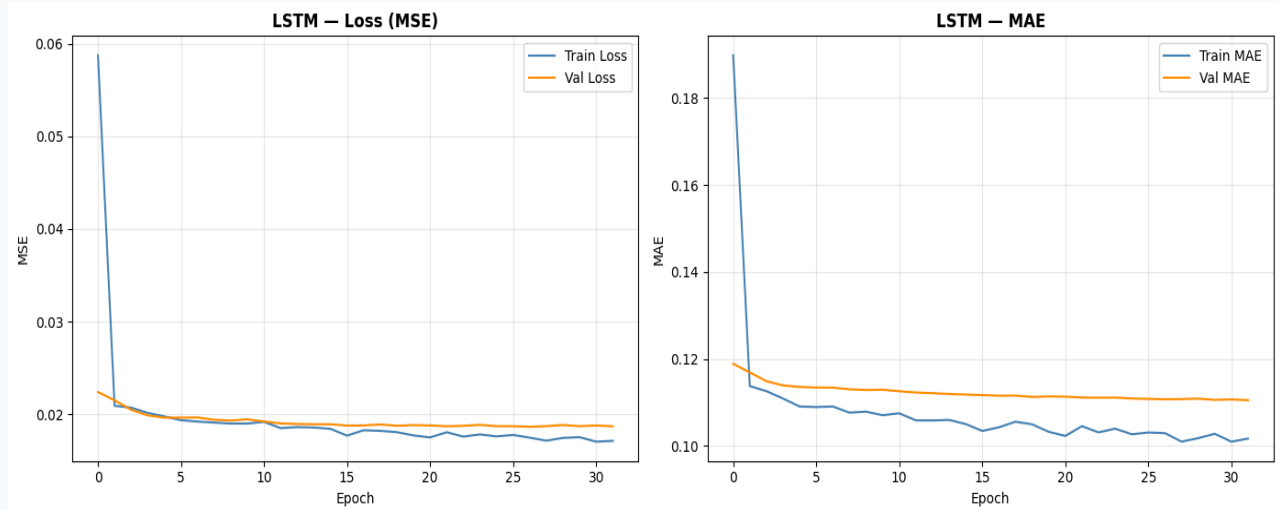
RESULTADOS — SIMPLERNN (w=90, n=32, LeakyReLU, bs=128)



MAPE test: 23,13% · MAE: 5,125 · RMSE: 6,505 (validacion: 25,12%)

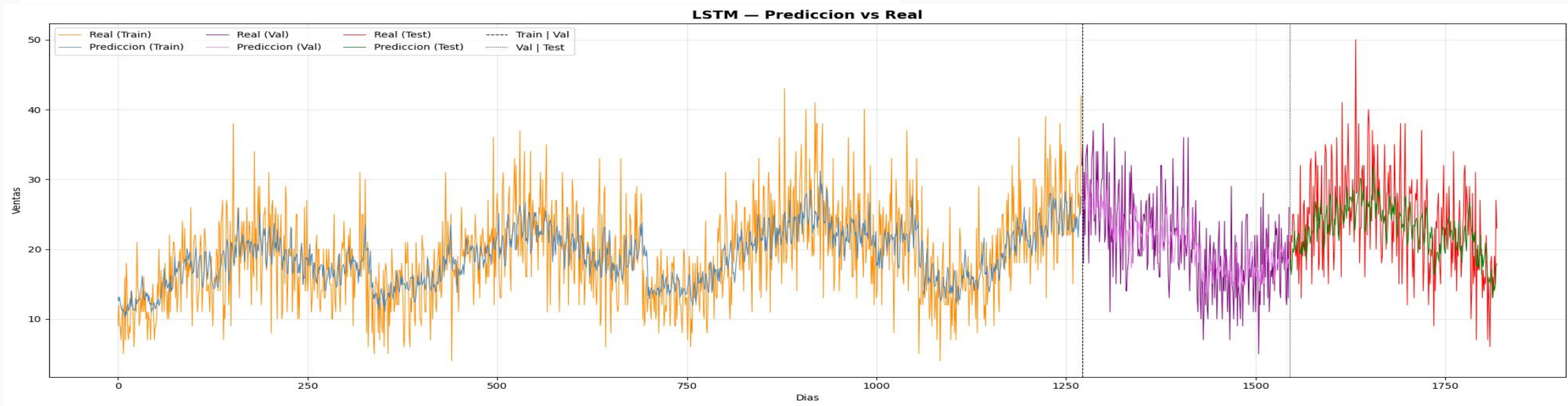
Solo 7 epochs: con lotes de 128 da pocas actualizaciones y el EarlyStopping corta pronto. Es la unica que NO supera al baseline en MAE ni en RMSE.



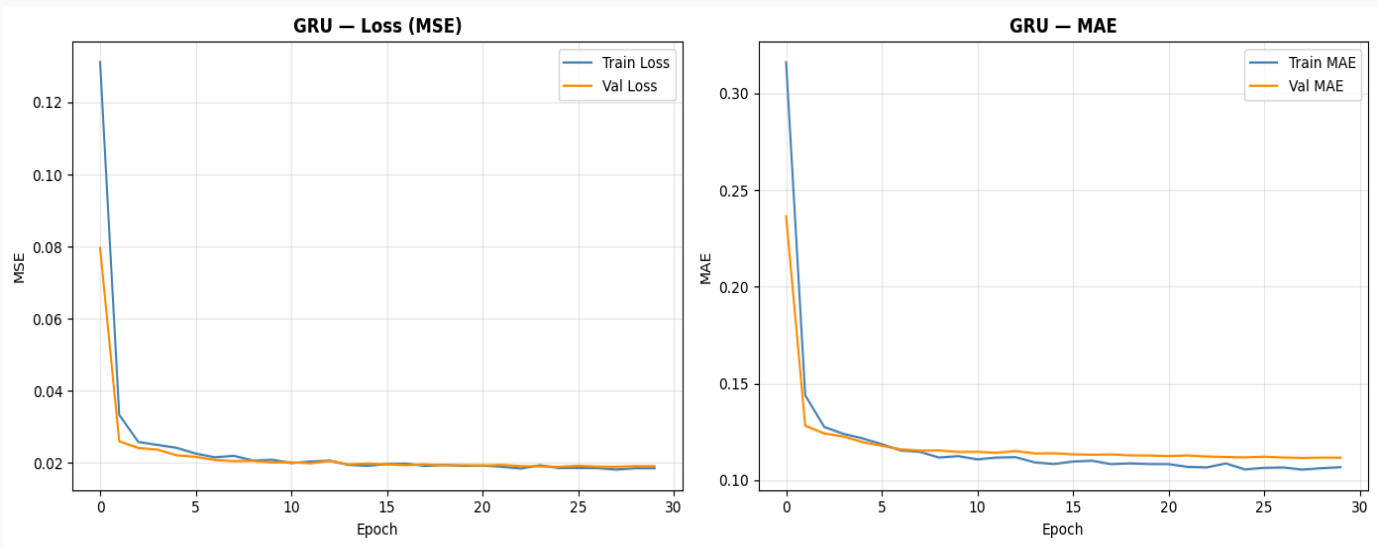


MAPE test: 20,95% · MAE: 4,550 · RMSE: 5,781 (validacion: 23,63%)

32 epochs, mejor punto en la 27. El val_loss mas bajo de los cuatro. Gana en las tres metricas de test.

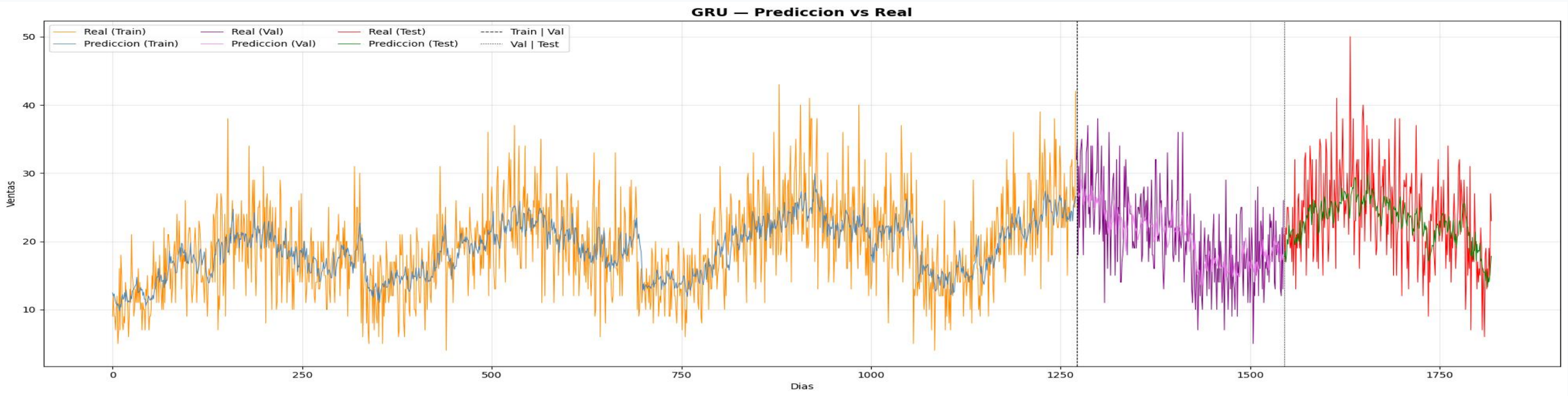


RESULTADOS — GRU (w=7, n=32, relu, bs=32)

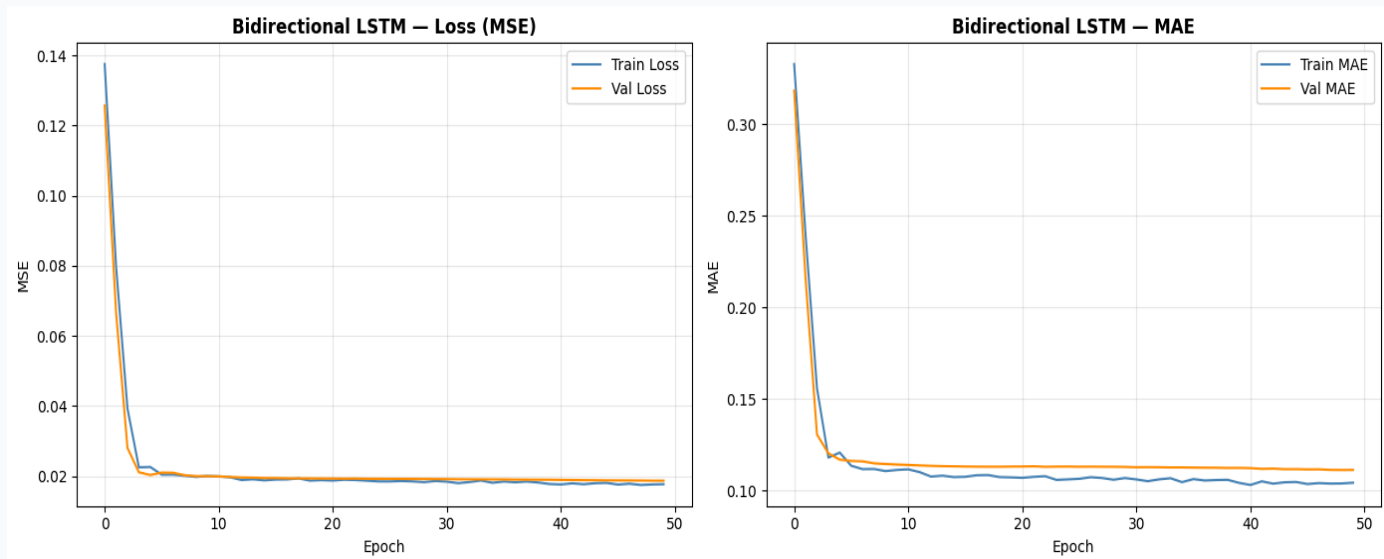


MAPE test: 21,42% · MAE: 4,622 · RMSE: 5,887
(validacion: 23,79%)

30 epochs, mejor punto en la 25. Supera al baseline en las tres metricas.



RESULTADOS — BI-LSTM (w=7, n=32, relu, bs=128)



MAPE test: 21,70% · MAE: 4,587 · RMSE: 5,811
(validacion: 23,96%)

50 epochs: llego al tope con su mejor punto en la 49, asi que seguia mejorando cuando se acabaron las epocas. Con mas presupuesto podria dar algo mas.

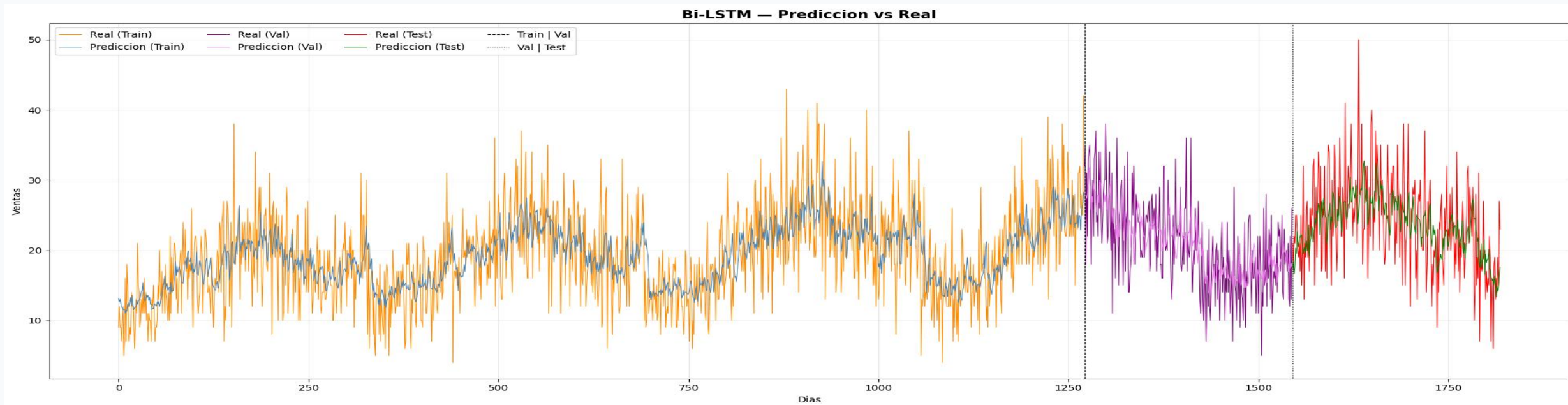


TABLA COMPARATIVA FINAL

Modelo	Config	MAPE Test	MAE Test	RMSE Test
SimpleRNN	w=90, n=32, act=LeakyReLU, bs=128	23,13%	5,125	6,505
LSTM	w=7, n=32, act=relu, bs=16	20,95%	4,550	5,781
GRU	w=7, n=32, act=relu, bs=32	21,42%	4,622	5,887
Bi-LSTM	w=7, n=32, act=relu, bs=128	21,70%	4,587	5,811
Media Movil	k=7	23,30%	4,867	6,124

Lectura honesta de la tabla:
Las TRES arquitecturas con puertas superan al baseline en las tres metricas.
La SimpleRNN NO: pierde en MAE (5,125 frente a 4,867) y en RMSE.
Entre LSTM y Bi-LSTM hay un 0,15% de la venta media: con una sola semilla por configuracion, eso no distingue nada.

1

Hay un grupo bueno, no un ganador

LSTM (4,55), Bi-LSTM (4,587) y GRU (4,622) superan al baseline en las tres metricas. Entre la primera y la segunda hay un 0,15% de la venta media: no basta para ordenarlas.

2

Las puertas sustituyen a la ventana larga

Las tres arquitecturas con gating eligieron ventana de 7 dias. La SimpleRNN, la unica sin ellas, necesito 90. Es lo que predice la teoria y es el resultado mas limpio del trabajo.

3

La SimpleRNN pierde contra la media movil

MAE 5,125 frente a 4,867 del baseline, y RMSE 6,505 frente a 6,124. La arquitectura recurrente mas simple no bate a una media de siete dias.

4

Si hay sobreajuste, y antes no se veia

En las 36 configuraciones el error de validacion es peor que el de train, sin excepcion. Antes se medía con MAPE y salia al revés: era un artefacto de escala, porque las ventas crecen con los años y el tramo final tiene ventas mas altas.

Limitaciones: una sola serie (Store 1, Item 1), una sola variable (ventas pasadas) y una sola semilla por configuracion. Falta el dia de la semana, que mueve las ventas un 53% (15,6 el lunes, 23,8 el domingo). Rejilla reducida en esta ejecucion; la completa esta en el notebook.

Gracias por su atención

¿Preguntas?
