

Evaluación modulo 2: Parte II

En este informe se presenta la implementación y comparación de los modelos MARS, LSTM, Random Forest (RF) y Singular Spectrum Analysis (SSA) aplicados a una serie de tiempo univariada correspondiente a los precios de cierre del activo META. El análisis abarca el periodo comprendido entre el 1 de mayo de 2016 y el 28 de enero de 2018. La evolución temporal de la serie se muestra en la Figura 1, donde la información fue descargada de Yahoo Finance.

Serie de tiempo de precios de cierre: META



Figura 1. Serie de tiempo de precios de cierre: Activo META. Se presenta la división de entrenamiento y prueba.

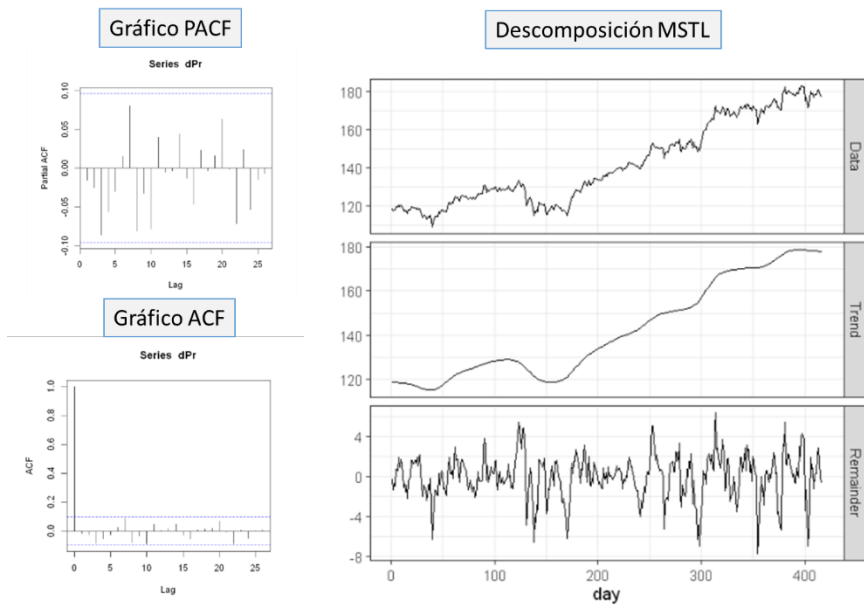


Figura 2. Gráficos PACF, ACF y descomposición con MSTL

una partición de los datos, asignando el 95% al conjunto de entrenamiento y el 5% restante al conjunto de prueba (Figura 1). Posteriormente, se llevó a cabo una exploración preliminar para identificar características relevantes de la serie temporal y posibles variables explicativas (features). Para ello, se utilizó la función `tsfeatures` junto con los gráficos de autocorrelación (ACF) y autocorrelación parcial (PACF), con el fin de evaluar la influencia de los rezagos (Figura 2).

Los resultados indican que la serie temporal de META presenta una fuerte tendencia (coeficiente de tendencia ≈ 0.99), alta autocorrelación ($x_acf1 \approx 0.99$) y baja entropía (0.093), lo que sugiere un comportamiento altamente predecible y determinista, aunque no estacionario en su

forma original. Además, la elevada linealidad (19.2) y una curvatura moderada (4.39) indican la presencia de patrones tanto lineales como no lineales, que pueden ser capturados por modelos suficientemente flexibles.

Aunque el PACF no señala rezagos específicos como predictores claves, los altos valores de autocorrelación en los primeros lags (x_acf1 y x_acf10) respaldan la inclusión de variables rezagadas simples (por ejemplo, `lag_1`, `lag_10`) como posibles entradas, especialmente cuando se trabaja sobre la serie original o suavizada en lugar de su versión diferenciada. Adicionalmente, con la descomposición MSTL (Multiple Seasonal-Trend decomposition using Loess) se observa que no hay un patrón estacional fuerte en la serie.

Modelos

Se implementaron y evaluaron cuatro modelos de pronóstico. En modelo **MARS** se evaluaron distintas combinaciones de rezagos y términos de tendencia cuadrática y cúbica, tanto de forma manual como mediante la librería ModelTime. Aunque la serie presenta una tendencia marcada, durante el proceso de modelado se observó que el modelo con mejor desempeño fue aquel que incluyó únicamente el rezago 1 como variable explicativa.

Para la red neuronal **LSTM** se definió previamente la estructura de datos transformados en un formato supervisado. Aunque el modelo mostró buenos resultados en términos de predicción, al analizar la evolución de la función de pérdida en el conjunto de entrenamiento se evidenció cierta inestabilidad, un comportamiento común en redes neuronales recurrentes. Posteriormente, el modelo fue incrustado nuevamente en la estructura de datos para realizar el pronóstico sobre el conjunto de prueba. Las métricas de desempeño obtenidas se presentan en la Figura 3.

En relación con los métodos de ensamble, se implementó el modelo **Random Forest**. Tras realizar el preprocesamiento y la partición, se construyó un modelo base que fue ajustado mediante la optimización de los hiperparámetros. De manera similar a MARS, se probaron diversas combinaciones de rezagos y configuraciones. El mejor desempeño se alcanzó con un modelo que incluía únicamente el rezago 1, una profundidad máxima de 10 y 100 árboles de decisión.

Finalmente, se implementó el modelo **SSA**, siguiendo una los pasos adecuados, y para la selección de los tripletes de vectores característicos, se utilizaron herramientas como la matriz de correlación-W, el scree plot y las gráficas de componentes en 1D y 2D. A partir de estos análisis, se propuso que los vectores 1, 2, 3, 4, 5 y 7 eran los candidatos ideales para representar la tendencia, mientras que los pares de vectores 8 y 9, y 10 y 11 se asociaron con la estacionalidad.

Resultados

Métricas de desempeño

Al comparar las métricas de desempeño (Figura 3), al evaluar las predicciones en el set de train, se observa que el modelo **LSTM** presenta la menor variabilidad en los errores de predicción (menor RMSE), así como los errores promedio más bajos (MAE). Este buen desempeño puede atribuirse a que, al tratarse de una red neuronal especializada en el manejo de secuencias, el modelo logra capturar relaciones temporales de largo plazo, así como patrones tanto lineales como no lineales complejos, que fueron evidentes al caracterizar la serie temporal. Aunque MARS no superó a LSTM, obtuvo resultados competitivos. Esto posiblemente se debe a su capacidad para modelar relaciones no lineales a través de *splines* adaptativos, lo cual le permitió ajustarse adecuadamente a los cambios presentes en la serie.

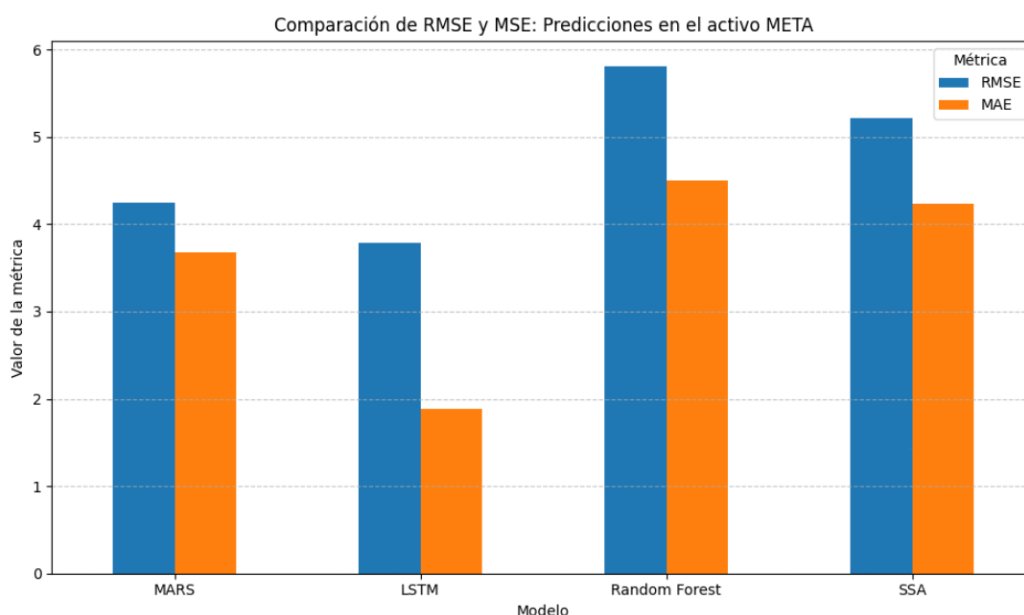


Figura 3. Comparación de métricas de desempeño modelos MARS, LSTM, Random Forest y SSA.

Pronóstico de 10 días hacia adelante

Se genera el pronóstico de 10 días hacia adelante utilizando el modelo de LSTM (Tabla 1 y Figura 4).

Tabla 1. Resultados: Pronósticos para los siguientes 10 días

Fecha	27/01/2018	28/01/2018	29/01/2018	30/01/2018	31/01/2018	1/02/2018	2/02/2018	3/02/2018	4/02/2018	5/02/2018
Valor Pronosticado	184.69	183.61	183.69	184.29	184.86	185.49	185.89	185.81	185.85	185.82

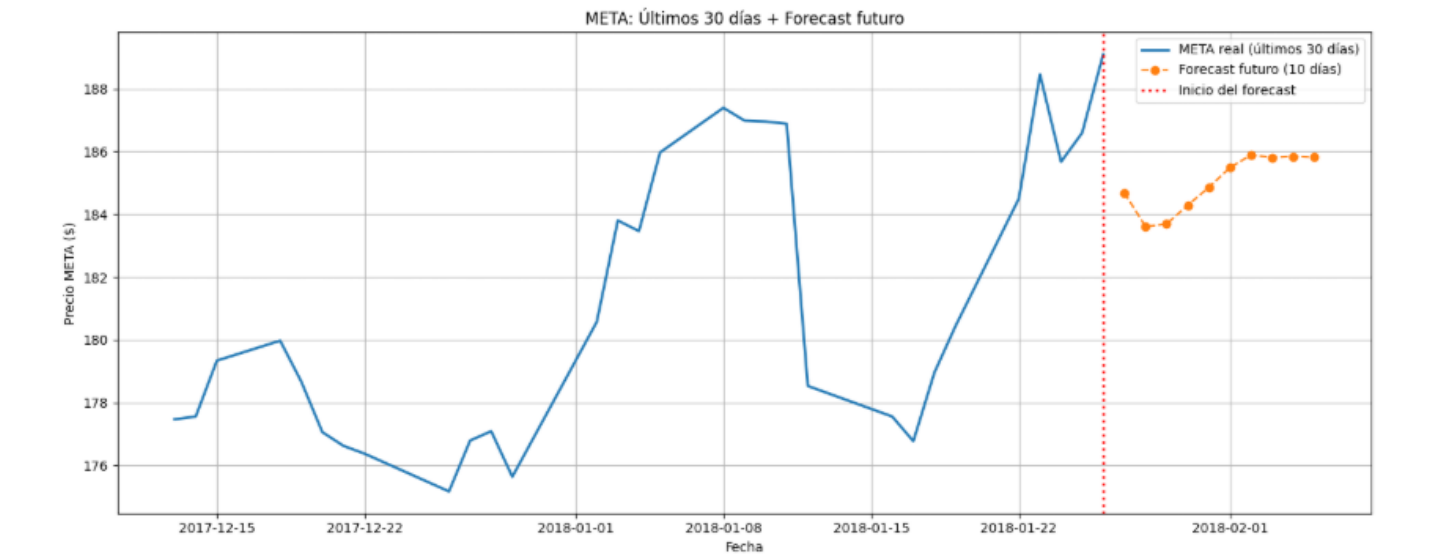


Figura 4. Visualización del pronóstico hacia adelante para la serie de META

Discusión

Cada uno de los modelos utilizados presenta ventajas y limitaciones particulares. El modelo LSTM demostró ser el más eficaz en términos de error de pronóstico, lo cual resalta su capacidad para capturar relaciones temporales de largo plazo y patrones complejos en series con comportamiento no lineal. Sin embargo, su entrenamiento puede ser computacionalmente costoso y sensible a la configuración de hiperparámetros, lo que lo hace menos interpretable y más difícil de ajustar sin experiencia técnica.

Por su parte, MARS ofreció un buen balance entre rendimiento y simplicidad. Su capacidad para identificar relaciones no lineales mediante splines adaptativos le permite capturar transiciones abruptas en la serie, siendo además más interpretable que LSTM. En contraste, Random Forest, aunque robusto ante el sobreajuste y útil para encontrar relaciones no lineales, presentó un peor desempeño con rezagos simples, posiblemente por su naturaleza de árbol que no modela explícitamente la temporalidad. Finalmente, SSA, si bien útil para descomponer la serie en componentes interpretables, depende fuertemente de una adecuada selección de vectores, y puede no ser tan competitivo como los modelos anteriores si no se optimiza esta etapa.

Conclusión

Se evidenció que los modelos con capacidad de capturar estructuras no lineales y dependencias temporales complejas, como LSTM y MARS, son más adecuados para el pronóstico de la serie META. LSTM se posiciona como el mejor modelo en este caso, gracias a su habilidad para aprender dinámicas temporales profundas, aunque su implementación y ajuste requieren mayores recursos y experiencia.

Link: Implementación de los modelos

https://github.com/ikteran/Pronostico_Mod2.git