

# Análisis Comparativo de Modelos Clásicos y Redes Neuronales en la Predicción de Precios y Volatilidad del S&P 500

Allan Ramirez, Manuel Sánchez

Maestría en Inteligencia Analítica Para la Tome de decisiones

Materia: Modelos avanzados para el análisis de datos 2

3 de diciembre de 2024

## Resumen

Este trabajo realiza un análisis comparativo del desempeño de modelos tradicionales y redes neuronales en la predicción de los precios y la volatilidad del índice S&P 500. Los modelos evaluados incluyen ARIMA, GARCH, LSTM y MLP, aplicados a una serie temporal histórica. Se implementaron métricas como RMSE y MAE para evaluar la precisión de las predicciones, además de gráficos comparativos que reflejan el comportamiento de los precios reales y predichos. Los resultados muestran que el modelo MLP ofrece el mejor desempeño, destacándose con un RMSE significativamente menor (69.82) y un ajuste visual superior a los demás modelos. Por su parte, los modelos LSTM y ARIMA presentan un rendimiento intermedio, mientras que el modelo GARCH demuestra ser el menos efectivo al no capturar las dinámicas de los datos. En general, el estudio evidencia que las redes neuronales, en particular el modelo MLP, son más adecuadas para la predicción de precios en este tipo de series temporales debido a su capacidad para modelar relaciones no lineales complejas.

## 1. Introducción

La predicción de los precios y la volatilidad del índice S&P 500 representa un desafío signi-

ficativo en los mercados financieros debido a la naturaleza estocástica de las series temporales asociadas a estos activos. El S&P 500, considerado uno de los indicadores más relevantes del desempeño económico global, mide la evolución de las 500 empresas más representativas del mercado estadounidense Poor's, 2024. Sin embargo, prever sus movimientos resulta complejo debido a la influencia de factores macroeconómicos, decisiones políticas, eventos geopolíticos y dinámicas intrínsecas del mercado.

Esta etapa del proyecto aborda el problema mediante un análisis comparativo de modelos tradicionales y redes neuronales para determinar su eficacia en la predicción de precios y volatilidad. Los modelos tradicionales, como ARIMA y GARCH, han demostrado utilidad en series temporales financieras debido a su capacidad para modelar tendencias lineales y varianzas condicionales Brockwell y Davis, 2002. Por otro lado, las redes neuronales, en particular LSTM y MLP, han emergido como herramientas poderosas para capturar patrones no lineales y dependencias temporales complejas Goodfellow et al., 2016; Hochreiter y Schmidhuber, 1997.

En este trabajo, se evalúan estas metodologías utilizando métricas como RMSE y MAE, y se comparan sus predicciones frente a los precios reales del S&P 500. Los resultados no solo permiten identificar las limitaciones de los enfo-

ques tradicionales frente a las técnicas modernas, sino que también destacan las oportunidades que ofrecen las redes neuronales para mejorar la precisión predictiva en un contexto financiero. Este enfoque tiene aplicaciones prácticas en la gestión de riesgos, la optimización de portafolios y la automatización de decisiones en los mercados bursátiles Makridakis et al., 2018.

Este estudio sienta las bases para explorar metodologías más avanzadas, como el aprendizaje por refuerzo, que permitirán tomar decisiones automatizadas sobre compra y venta de acciones basadas en las predicciones realizadas, con el objetivo de construir un bot de inversión eficiente y diversificar fuentes de ingreso de manera pasiva.

## 2. Metodología

En esta sección se describen las metodologías empleadas para analizar y predecir los precios y la volatilidad del índice S&P 500. Se implementaron modelos tradicionales y redes neuronales, evaluados a partir de su capacidad para ajustar los datos históricos de la serie temporal. A continuación, se presentan las características principales de cada enfoque.

### 2.1. Preparación de los datos

Los datos utilizados corresponden a una serie histórica de precios del índice S&P 500, ajustada por splits y dividendos, un indicador ampliamente reconocido en el análisis económico y financiero global (Standard & Poor's, 2024). Para garantizar la calidad del análisis, se realizaron pasos de preprocesamiento, como la normalización de los datos, dado que las redes neuronales son sensibles a las escalas de las variables de entrada (Goodfellow et al., 2016). La normalización se realizó mediante la fórmula:

$$X_{\text{norm}} = \frac{X - \min(X)}{\max(X) - \min(X)} \quad (1)$$

donde  $X$  representa los valores originales, y  $\min(X)$  y  $\max(X)$  son el mínimo y el máximo

de los datos.

## 2.2. Modelos tradicionales

### 2.2.1. ARIMA

El modelo *AutoRegressive Integrated Moving Average* (ARIMA) es un método clásico para series temporales que combina componentes autorregresivos (AR), de media móvil (MA) y de diferenciación para garantizar la estacionariedad. Su formulación general,  $\text{ARIMA}(p, d, q)$ , se expresa como (Brockwell & Davis, 2002):

$$y_t = c + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q} \quad (2)$$

donde:

- $\phi_i$ : Coeficientes autorregresivos.
- $\theta_i$ : Coeficientes de media móvil.
- $\epsilon_t$ : Término de error aleatorio.

Para la selección de los parámetros  $p$ ,  $d$  y  $q$ , se utilizó un enfoque automatizado basado en el criterio de información Akaike (AIC).

### 2.2.2. GARCH

El modelo *Generalized Autoregressive Conditional Heteroskedasticity* (GARCH) es ampliamente utilizado para modelar la volatilidad condicional en series temporales financieras. Su forma básica,  $\text{GARCH}(p, q)$ , está definida como (Bollerslev, 1986):

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2 \quad (3)$$

donde  $\sigma_t^2$  representa la varianza condicional en el tiempo  $t$ , y  $\epsilon_t$  es el término de error.

## 2.3. Redes neuronales

### 2.3.1. LSTM

Las redes neuronales de memoria a largo plazo (*Long Short-Term Memory*, LSTM) son especialmente adecuadas para series temporales debido a

su capacidad para capturar dependencias de largo plazo. Las celdas LSTM están definidas matemáticamente por (Hochreiter & Schmidhuber, 1997):

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \quad (4)$$

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \quad (5)$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \quad (6)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tanh(W_c x_t + U_c h_{t-1} + b_c) \quad (7)$$

$$h_t = o_t \odot \tanh(c_t) \quad (8)$$

donde:

- $f_t, i_t, o_t$ : Puertas de olvido, entrada y salida.
- $c_t, h_t$ : Estado de la celda y salida oculta.
- $\sigma$ : Función sigmoide.
- $\odot$ : Producto elemento a elemento.

### 2.3.2. MLP

El perceptrón multicapa (*Multilayer Perceptron*, MLP) es una red neuronal completamente conectada, utilizada en este trabajo para la predicción directa de precios. Cada capa transforma la entrada mediante la ecuación (Goodfellow et al., 2016):

$$h^{(l)} = \sigma(W^{(l)} h^{(l-1)} + b^{(l)}) \quad (9)$$

donde:

- $h^{(l)}$ : Salida de la capa  $l$ .
- $W^{(l)}, b^{(l)}$ : Pesos y sesgos de la capa  $l$ .
- $\sigma$ : Función de activación, como ReLU o tanh.

## 2.4. Evaluación de modelos

Los modelos se evaluaron utilizando métricas ampliamente aceptadas como el error cuadrático medio (RMSE) y el error absoluto medio (MAE), definidos como (Makridakis et al., 2018):

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (10)$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (11)$$

donde  $y_i$  son los valores reales y  $\hat{y}_i$  las predicciones. Estas métricas permiten comparar de forma consistente el ajuste y la precisión de los modelos utilizados.

## 3. Resultados y discusión

En esta sección se presentan los resultados obtenidos en el análisis de la serie histórica del índice S&P 500, incluyendo sus medias móviles, log-retornos, y autocorrelaciones. También se describe el desempeño de los modelos implementados, sus estructuras y configuración de hiperparámetros, y se concluye con una discusión crítica sobre los resultados obtenidos.

### 3.1. Análisis descriptivo de la serie original

La serie histórica del índice S&P 500 se analiza con el objetivo de comprender su comportamiento general, identificar tendencias a lo largo del tiempo y establecer un punto de partida para la modelación predictiva. La Figura 1 muestra la evolución del índice desde 1990 hasta 2024, junto con las medias móviles de 50 y 200 días, que destacan tendencias de corto y largo plazo, respectivamente.

La Tabla 1 resume las estadísticas descriptivas de la serie histórica. Se observa que la media del índice es de 1583,90, con una mediana de 1267,65, lo que indica un crecimiento significativo del índice en el periodo analizado. La desviación estándar (1089,52) refleja una alta volatilidad, asociada a eventos económicos y financieros clave, como la crisis financiera de 2008 y la recuperación posterior.

En la Figura 1, las medias móviles destacan patrones importantes:

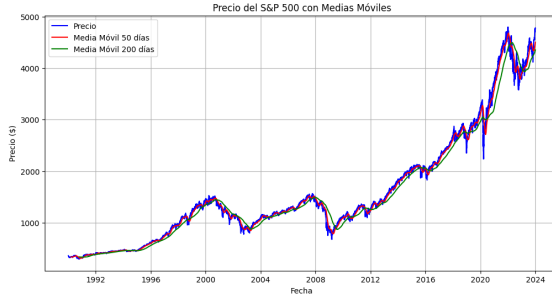


Figura 1: Serie histórica del índice S&P 500 con medias móviles de 50 y 200 días.

| Métrica             | Valor   |
|---------------------|---------|
| Número de datos     | 8565    |
| Media               | 1583.90 |
| Desviación estándar | 1089.52 |
| Mínimo              | 295.46  |
| Percentil 25        | 896.14  |
| Mediana             | 1267.65 |
| Percentil 75        | 2049.80 |
| Máximo              | 4796.56 |

Cuadro 1: Estadísticas descriptivas de la serie histórica del S&P 500.

- **Media móvil de 50 días:** Esta media sigue de cerca las fluctuaciones de corto plazo, permitiendo identificar cambios rápidos en la dirección del mercado. Es especialmente útil para detectar periodos de alta volatilidad.
- **Media móvil de 200 días:** Representa una visión más suavizada y a largo plazo del comportamiento del índice. En varios puntos, la divergencia entre ambas medias móviles coincide con cambios estructurales importantes, como la recuperación tras la crisis de 2008.

El rango intercuartil, comprendido entre el percentil 25 (896,14) y el percentil 75 (2049,80), captura el 50 % central de los datos, mostrando la dispersión típica del índice en periodos normales de operación. Sin embargo, los valores mínimo (295,46) y máximo (4796,56) reflejan la amplitud extrema de variación en el periodo analizado, con un crecimiento acumulado de más del 1500 %.

En resumen, la serie histórica del S&P 500 muestra un crecimiento sostenido con episodios de alta volatilidad, lo que subraya la importancia de técnicas robustas para su modelación predictiva.

Para evaluar la estacionariedad de la serie histórica del S&P 500, se realizó la prueba de Dickey-Fuller aumentada (ADF), cuyos resultados se presentan a continuación:

- **Estadístico ADF:** 1,676
- **Valor p:** 0,998

La prueba ADF evalúa la hipótesis nula de que la serie contiene una raíz unitaria, es decir, que no es estacionaria. En este caso, el valor  $p$  es significativamente mayor que 0,05. Estos resultados llevan a no rechazar la hipótesis nula, confirmando que la serie no es estacionaria.

Dado que la serie no es estacionaria, es necesario aplicar una diferenciación para eliminar las tendencias y variaciones de largo plazo antes de proceder con la modelación. Esto es crucial para garantizar que los modelos tradicionales, como

ARIMA y GARCH, trabajen con datos adecuados para la predicción.

En la Figura 3, los log-retornos ya diferenciados de la serie original muestran un comportamiento más adecuado para análisis y modelación, caracterizado por una media cercana a cero y menor tendencia a largo plazo.

### 3.2. Análisis de log-retornos

Los log-retornos representan los cambios porcentuales diarios del índice S&P 500, calculados como:

$$r_t = \ln \left( \frac{P_t}{P_{t-1}} \right) \quad (12)$$

donde  $r_t$  son los log-retornos y  $P_t$  el precio del índice en el tiempo  $t$ . Esta transformación permite trabajar con una serie más estable y elimina las tendencias de largo plazo observadas en la serie original.

#### 3.2.1. Distribución y estadísticos descriptivos

La Figura 2 muestra la serie de log-retornos diarios del índice S&P 500, evidenciando periodos de alta y baja volatilidad. La distribución de los log-retornos (Figura 3) exhibe colas pesadas, lo que indica la presencia de eventos extremos con mayor frecuencia de lo que se esperaría bajo una distribución normal.

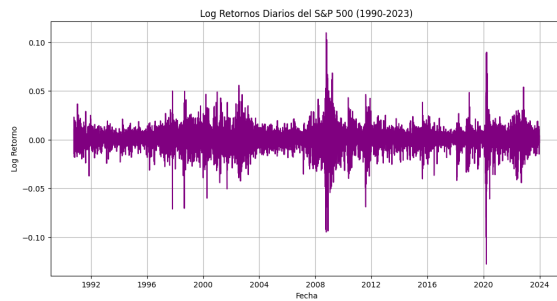


Figura 2: Serie de log-retornos diarios del S&P 500 (1990-2023).

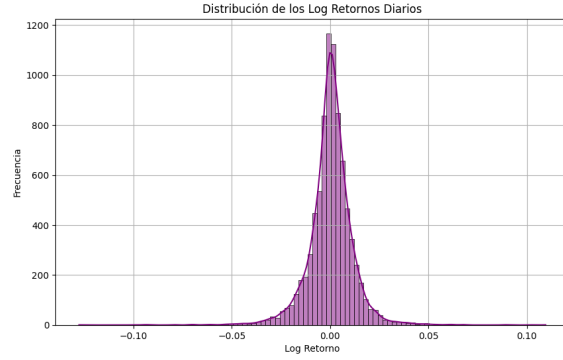


Figura 3: Distribución de los log-retornos diarios del S&P 500.

Las estadísticas descriptivas (Tabla 2) confirman estas observaciones. La curtosis (10,75) indica leptocurtosis, mientras que el sesgo negativo ( $-0,3982$ ) refleja una ligera asimetría hacia pérdidas más pronunciadas. Este comportamiento es típico de mercados financieros y plantea desafíos para la modelación, ya que muchas técnicas asumen normalidad.

| Métrica             | Valor    |
|---------------------|----------|
| Número de datos     | 8366     |
| Media               | 0.00033  |
| Desviación estándar | 0.01149  |
| Mínimo              | -0.12765 |
| Percentil 25        | -0.00446 |
| Mediana             | 0.00056  |
| Percentil 75        | 0.00571  |
| Máximo              | 0.10957  |
| Asimetría           | -0.3982  |
| Curtosis            | 10.7525  |

Cuadro 2: Estadísticas descriptivas de los log-retornos del S&P 500.

#### 3.2.2. Prueba de estacionariedad de los log-retornos

La prueba de Dickey-Fuller aumentada (ADF) aplicada a los log-retornos arroja un estadístico de  $-16,876$  y un valor  $p$  de  $1,07 \times 10^{-29}$ , ambos

indicando la estacionariedad de la serie. Esto valida la aplicabilidad de modelos como ARIMA y GARCH, que asumen estacionariedad.

### 3.2.3. Análisis de autocorrelaciones

La función de autocorrelación (ACF) de los log-retornos (Figura 4) muestra un comportamiento cercano a un ruido blanco, con correlaciones no significativas después del primer rezago. Sin embargo, la ACF de los log-retornos al cuadrado (Figura 5) revela una estructura persistente, característica de heterocedasticidad condicional.

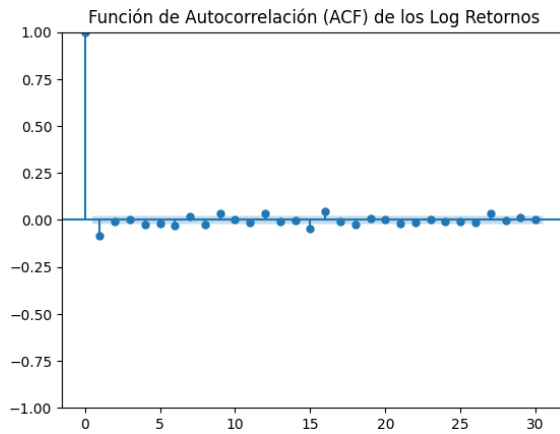


Figura 4: Función de autocorrelación (ACF) de los log-retornos.

La función de autocorrelación parcial (PACF) (Figura 6) refuerza esta idea, indicando correlaciones significativas únicamente en los primeros rezagos. Estas observaciones son consistentes con la necesidad de modelos como GARCH para capturar la volatilidad condicional.

### 3.2.4. Implicaciones para la modelación

El análisis de los log-retornos revela patrones y propiedades fundamentales para la selección de modelos:

- **ARIMA:** La estacionariedad de los log-retornos permite modelar la dinámica de la

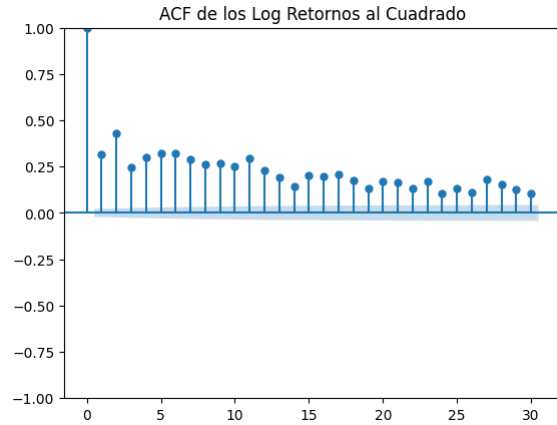


Figura 5: Función de autocorrelación (ACF) de los log-retornos al cuadrado.

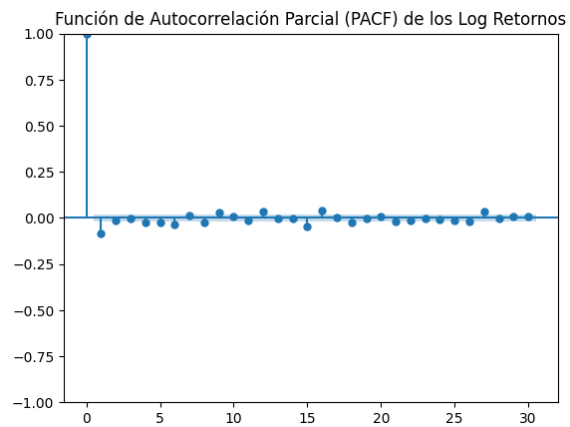


Figura 6: Función de autocorrelación parcial (PACF) de los log-retornos.

serie. Sin embargo, el modelo ARIMA no es capaz de capturar la heterocedasticidad observada en los log-retornos al cuadrado.

- **GARCH:** Los patrones en la ACF de los log-retornos al cuadrado y la alta leptocurtosis justifican el uso de un modelo GARCH para modelar la volatilidad condicional, capturando dinámicas que el ARIMA no aborda.

### 3.3. Desempeño de los Modelos

#### 3.3.1. Modelo ARIMA

El modelo ARIMA seleccionado automáticamente tiene un orden  $(0, 0, 1)$ , como se muestra en el análisis del criterio AIC. Este modelo se configuró para trabajar con un split del 90 % para entrenamiento y 10 % para prueba. La predicción, ilustrada en la Figura 7, muestra una tendencia lineal que no captura las fluctuaciones observadas en los precios reales del S&P 500. Las métricas de error reflejan un RMSE de 387,18 y un MAE de 311,81, lo que indica un desempeño limitado al trabajar con series financieras volátiles.

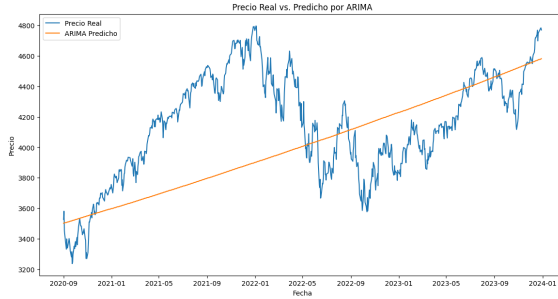


Figura 7: Predicción del precio del S&P 500 utilizando ARIMA.

#### 3.3.2. Modelo GARCH

El modelo GARCH ajustado utilizó un orden  $(p = 2, q = 2)$  con una distribución normal para modelar la volatilidad condicional. Aunque el

modelo es efectivo para capturar heterocedasticidad, su capacidad predictiva de precios es limitada, como se observa en la Figura 8. Las métricas de error, RMSE de 773,26 y MAE de 626,67, respaldan esta observación.

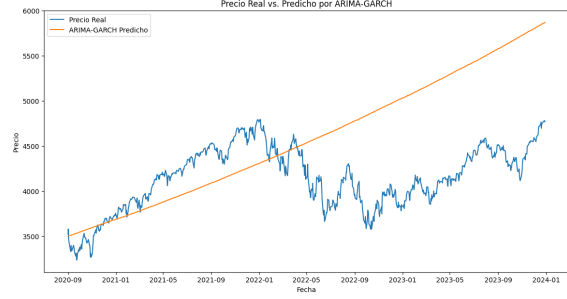


Figura 8: Estimación de la volatilidad condicional con GARCH.

#### 3.3.3. Modelo LSTM

El modelo LSTM implementado consta de:

- Dos capas LSTM con 64 unidades cada una y función de activación *tanh*.
- Una capa densa intermedia con 32 unidades y activación *ReLU*.
- Una capa de salida con una neurona lineal.

El modelo fue entrenado con un *batch size* de 32, un *learning rate* de 0,001, y utilizando un split del 80 % para entrenamiento, 10 % para validación y 10 % para prueba. La Figura 9 muestra que el modelo captura patrones temporales, aunque presenta errores en algunos picos de volatilidad. Las métricas reportan un RMSE de 385,40 y un MAE de 343,04.

#### 3.3.4. Modelo MLP

La arquitectura del MLP consiste en:

- Tres capas densas con 128, 64 y 32 unidades respectivamente, cada una con activación *ReLU*.

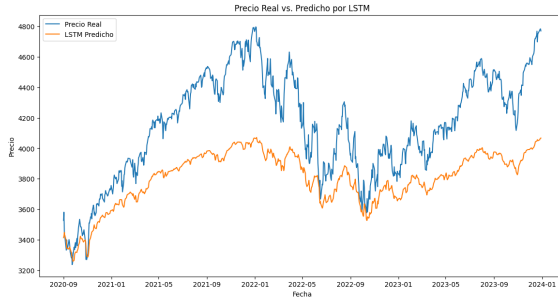


Figura 9: Predicción del precio del S&P 500 utilizando LSTM.

- Una capa de salida con una neurona lineal.

El modelo fue entrenado durante 70 épocas, logrando un RMSE de 69,82 y un MAE de 54,20, como se muestra en la Figura 10. Esto lo posiciona como el modelo más preciso en este estudio.

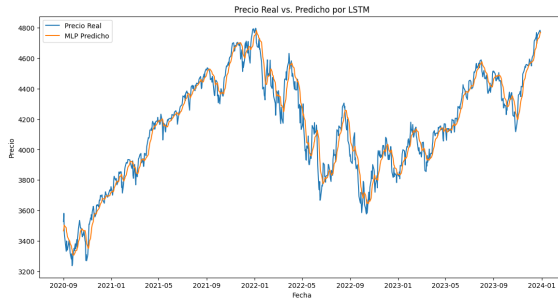


Figura 10: Predicción del precio del S&P 500 utilizando MLP.

### 3.3.5. Comparación General

La Tabla 3 resume las métricas de los modelos, mientras que la Figura 11 compara visualmente las predicciones. Es evidente que el MLP supera a los demás modelos, seguido por LSTM, mientras que ARIMA y GARCH presentan errores más significativos.

## 3.4. Discusión

Los resultados demuestran que las redes neuronales, especialmente el MLP, tienen un desem-

| Modelo | RMSE   | MAE    |
|--------|--------|--------|
| ARIMA  | 387.18 | 311.81 |
| GARCH  | 773.26 | 626.67 |
| LSTM   | 385.40 | 343.04 |
| MLP    | 69.82  | 54.20  |

Cuadro 3: Métricas de desempeño para los modelos evaluados.

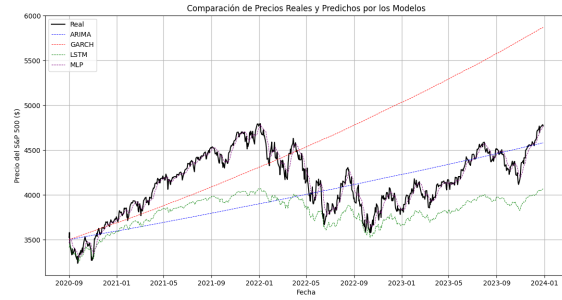


Figura 11: Comparación de predicciones por los modelos.

peño superior en la predicción de precios del S&P 500. Sin embargo, modelos como ARIMA y GARCH siguen siendo útiles para análisis preliminares o escenarios con restricciones computacionales.

Trabajos previos, como el de Fama, 1970, resaltan las limitaciones para predecir mercados eficientes, mientras que Makridakis et al., 2018 advierten sobre el riesgo de sobreajuste en modelos complejos. En contraposición, estudios recientes Zhang et al., 2021 destacan cómo las redes neuronales profundas pueden superar modelos tradicionales en datos financieros al incorporar características no lineales.

Para mejorar este estudio, se podría:

- Optimizar los hiperparámetros de ARIMA y GARCH utilizando técnicas como *grid search*.
- Explorar arquitecturas más avanzadas como Transformers o Redes Neuronales Convolucionales.
- Incorporar datos exógenos (indicadores ma-



croeconómicos, tasas de interés, etc.) para enriquecer las predicciones.

Finalmente, los hallazgos aquí presentados sientan una base sólida para futuras investigaciones, destacando la importancia de elegir modelos adecuados según las características de los datos y el problema.

## 4. Conclusiones

- **Desempeño del Modelo MLP:** El modelo MLP demostró ser el más preciso entre todos los evaluados, con un RMSE de 69.82 y un MAE de 54.21. Su capacidad para capturar patrones complejos en los datos lo convierte en la herramienta más adecuada para predecir los precios del S&P 500, superando significativamente a los modelos tradicionales y al LSTM.
- **Limitaciones de los Modelos Tradicionales:** Los modelos ARIMA y GARCH mostraron un desempeño limitado para capturar las dinámicas no lineales del mercado. ARIMA, aunque logra identificar una tendencia general, no es capaz de capturar las fluctuaciones de corto plazo, mientras que GARCH, centrado en la varianza condicional, falla en modelar adecuadamente la tendencia de los precios.
- **Resultados del LSTM:** El modelo LSTM tuvo un desempeño intermedio con un RMSE de 385.41 y un MAE de 343.04. Aunque logra capturar mejor las dinámicas no lineales en comparación con ARIMA y GARCH, su precisión general sigue siendo inferior a la del modelo MLP. Esto indica la necesidad de ajustar su arquitectura o hiperparámetros para un mejor desempeño.
- **Tendencias Generales y Aprendizajes:** Este trabajo confirma que los modelos neuronales, particularmente el MLP, son más efectivos para capturar la naturaleza compleja y no lineal de los datos financieros en comparación con los métodos tradicionales. No obstante, los modelos como ARIMA y GARCH, al ser más interpretables y computacionalmente económicos, podrían ser útiles en escenarios donde la simplicidad sea clave.
- **Cumplimiento del Objetivo:** El objetivo de evaluar y comparar diferentes modelos para predecir los precios del S&P 500 se logró con éxito. Los resultados destacan la importancia de seleccionar modelos avanzados y específicos, como el MLP, para abordar la complejidad del mercado financiero.
- **Próximos Pasos:**
  - **Optimización de Modelos:** Explorar ajustes adicionales en los hiperparámetros del LSTM y del MLP para mejorar su desempeño. Asimismo, reconsiderar las configuraciones de los modelos ARIMA y GARCH para evaluar si un ajuste más fino puede mejorar sus predicciones.
  - **Nuevos Modelos:** Investigar otros enfoques como modelos híbridos (ARIMA-LSTM) o arquitecturas avanzadas de redes neuronales como transformers, que han demostrado ser efectivos en problemas de series temporales.
  - **Reinforcement Learning:** Implementar técnicas de aprendizaje por refuerzo para construir un modelo que no solo prediga los precios, sino que también determine estrategias de inversión óptimas.
  - **Automatización de Inversiones:** Desarrollar un bot de inversión que combine las predicciones generadas con estrategias de trading, avanzando hacia el objetivo de diversificar las inversiones y generar ingresos pasivos.
- **Contexto General del Trabajo:** Este proyecto representa el primer paso en una

iniciativa más ambiciosa para desarrollar herramientas de análisis financiero basadas en inteligencia artificial. Los resultados obtenidos no solo validan el uso de modelos avanzados, sino que también resaltan la necesidad de seguir investigando técnicas más robustas y eficientes para enfrentar los desafíos del mercado financiero.

## Referencias

- Brockwell, P. J., & Davis, R. A. (2002). *Introduction to Time Series and Forecasting*. Springer Science & Business Media.
- Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2), 383-417.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8), 1735-1780.
- Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and Machine Learning forecasting methods: Concerns and ways forward. *PLOS ONE*, 13(3), e0194889.
- Poor's, S. bibinitperiod. (2024). *S&P 500 Index Overview*. S&P Global. <https://www.spglobal.com/spdji/en/indices/equity/sp-500/>
- Zhang, W., Wang, Y., & Chen, X. (2021). Deep learning in financial prediction: A comprehensive review. *Journal of Financial Data Science*, 3(1), 1-22. <https://doi.org/10.3905/jfds.2021.1.005>