Pontificia Universidad
JAVERIANA
Cali

PREDICCIÓN DE VOLATILIDAD EN SERIES DE TIEMPO FINANCIERAS CON MODELOS AUTORREGRESIVOS EN ESPACIOS DE HILBERT

Ginna Paola Quintero Afanador
Jesús David Cárdenas Barreto

Proyecto Aplicado para optar al título de
Magister en Ciencia de Datos

Director(a)
Cristhian Kaori Valencia Marín

FACULTAD DE INGENIERÍA Y CIENCIAS
MAESTRÍA EN CIENCIA DE DATOS
SANTIAGO DE CALI, DICIEMBRE 1 DE 2024

TABLA DE CONTENIDO

INTRODUCCIÓN 5
I. DEFINICIÓN DEL PROBLEMA 6
A. PLANTEAMIENTO DEL PROBLEMA 6
B. FORMULACIÓN DEL PROBLEMA 7
II. OBJETIVOS DEL PROYECTO 8
A. OBJETIVO GENERAL 8
B. OBJETIVOS ESPECÍFICOS 8
III. MARCO TEÓRICO Y ANTECEDENTES 9
A. MARCO TEÓRICO 9
1. SERIES DE TIEMPO FINANCIERAS 9
2. VOLATILIDAD 10
3. MODELOS AUTORREGRESIVOS 10
4. MODELO AR 10
5. SUAVIZAMIENTO CON MEDIA MÓVIL EXPONENCIALMENTE PONDERADA (EWMA) 11
6. MODELO ARCH 12
7. MODELO GARCH 13
8. ESPACIOS VECTORIALES 14
9. ESPACIOS DE HILBERT 14
10. TEOREMA DE MERCER 16
11. TEOREMA DE RIESZ 16
12. REDES NEURONALES 17
13. MODELOS DE BOSQUES ALEATORIOS 18
B. ANTECEDENTES 18
IV. METODOLOGÍA PARA EL PREPROCESAMIENTO DE UNA SERIE DE TIEMPO FINANCIERA 22
A. EXPLORACIÓN DE FUENTES Y SERIES DE TIEMPO FINANCIERAS 22
B. METODOLOGÍA PARA EL PREPROCESAMIENTO DE SERIES DE TIEMPO FINANCIERAS 26
1. LIMPIEZA DE DATOS 26
2. TRANSFORMACIÓN DE LOS DATOS 27
3. MANEJO DE LA ESTACIONALIDAD 27
4. VALIDACIÓN DE DATOS 27
V. MODELOS AUTORREGRESIVOS Y ESPACIOS DE HILBERT 31
A. ESPACIO DE HILBERT CON KERNEL REPRODUCTIVO 31
B. TIPOS DE KERNELS REPRODUCTIVOS 31
VI. APLICACIÓN DE MODELOS AUTORREGRESIVOS A SERIES DE TIEMPO FINANCIERAS 34
A. SERIE APPLE (APPL) 34
B. SERIE AMAZON (AMZN) 39
C. SERIE BOLSA DE VALORES DE ESTAMBUL (BVE) 45

VII. APLICACIÓN DE KERNELS REPRODUCTIVOS EN MODELOS AUTOREGRESIVOS 52
A. APLICACIÓN DE KERNELS REPRODUCTIVOS A MODELOS EN SERIE APPLE 52
B. APLICACIÓN DE KERNELS REPRODUCTIVOS A MODELOS EN SERIE AMAZON 53
C. APLICACIÓN DE KERNELS REPRODUCTIVOS A MODELOS EN SERIE BVE (SP) 55
VIII. APLICACIÓN DE MODELOS EN ESPACIOS DE HILBERT CON KERNEL REPRODUCTIVO 57
A. APLICACIÓN DE MODELOS EN ESPACIOS DE HILBERT PARA LA SERIE APPLE 57
B. APLICACIÓN DE MODELOS EN ESPACIOS DE HILBERT PARA LA SERIE AMAZON 58
C. APLICACIÓN DE MODELOS EN ESPACIOS DE HILBERT PARA LA SERIE BVE (SP) 59
IX. EVALUACIÓN DE MODELOS PARA LA PREDICCIÓN DE LA VOLATILIDAD 62
A. ANÁLISIS DE LOS RESULTADOS OBTENIDOS 64
1. ANÁLISIS DE LOS RESULTADOS OBTENIDOS PARA LA SERIE APPLE 64
2. ANÁLISIS DE LOS RESULTADOS OBTENIDOS PARA LA SERIE AMAZON 65
3. ANÁLISIS DE LOS RESULTADOS OBTENIDOS PARA LA SERIE BVE (SP) 66
X. CONCLUSIONES Y TRABAJOS FUTUROS 68
A. CONCLUSIONES 68
B. TRABAJOS FUTUROS 70
XI. REFERENCIAS BIBLIOGRÁFICAS 71

LISTA DE FIGURAS

Figura 1 Co-ocurrencia de palabras clave en bibliografía relacionada 2014-2022. Construcción propia.....19
Figura 2 Matriz de correlación BVE. Construcción propia. Fuente: Serie de tiempo BVE.....24
Figura 3 Matriz de correlación de la acción Amazon (AMZN). Construcción propia. Fuente: Serie de tiempo AMZN 25
Figura 4 Distribución de precios sobre la acción APPL. Construcción propia. Fuente: Serie de tiempo APPL. 26
Figura 5 Serie de tiempo de rendimientos de la Bolsa de Estambul. Construcción propia. Fuente: Serie de tiempo BVE. 28
Figura 6 Serie de tiempo de precios de la acción Amazon (AMZN). Construcción propia. Fuente: Serie de tiempo AMZN.....29
Figura 7 Análsis de retardo de la serie de tiempo Amazon (AMZN) Construcción propia. Fuente: Serie de tiempo AMZN. 30
Figura 8 Precios de la acción APPL. Construcción propia. Fuente: Serie de tiempo APPL. 30
Figura 9 Resultados predicción de la volatilidad del precio de la acción APPL con modelo AR. Construcción propia. 35
Figura 10 Gráfico de autocorrelación del modelo AR. Construcción propia. 35
Figura 11 Predicción del modelo AR con suavizamiento EWMA para la acción APPL. Construcción propia. 36
Figura 12 Predicciones del Modelo ARCH para las Acciones de Apple. Construcción propia. 37
Figura 13 Predicciones del Modelo ARCH bajo EWMA para las Acciones de Apple. Construcción propia. 38
Figura 14 Predicciones del Modelo GARCH para las Acciones de Apple. Construcción propia. 38
Figura 15 Predicciones del Modelo GARCH bajo EWMA para las Acciones de Apple. Construcción propia. 39
Figura 16 Predicción del modelo AR para la acción AMZN. Construcción propia. 40
Figura 17 Predicción del modelo AR suavizado con EWMA para la acción AMZN. Construcción propia. 41
Figura 18 Predicción del modelo ARCH para la acción AMZN. Construcción propia. 42
Figura 19 Predicción del modelo ARCH suavizado con EWMA para la acción AMZN. Construcción propia. 43
Figura 20 Predicción del modelo GARCH para la acción AMZN. Construcción propia. 44
Figura 21 Predicción del modelo GARCH suavizado con EWMA para la acción AMZN. Construcción propia. 45
Figura 22 Predicciones del Modelo AR para el índice SP. Construcción propia. 46
Figura 23 Predicciones del Modelo AR bajo EWMA para el índice SP. Construcción propia. 47
Figura 24 Predicciones del Modelo ARCH para el índice SP. Construcción propia. 48
Figura 25 Predicciones del Modelo ARCH bajo EWMA para el índice SP. Construcción propia. 48
Figura 26 Predicciones del Modelo GARCH para el índice SP. Construcción propia. 49
Figura 27 Predicciones del Modelo GARCH bajo EWMA para el índice SP. Construcción propia. 50
Figura 28 Comparación resultado gráfico de la aplicación de Kernels reproductivos a la serie AMZN. Construcción propia. 54
Figura 29 Comparación gráfica de resultados con la aplicación de espacios de Hilbert en la serie APPL. Construcción propia. 58
Figura 30 Comparación gráfica de resultados con la aplicación de espacios de Hilbert en la serie AMZN. Construcción propia. .59
Figura 31 Comparación gráfica de resultados con la aplicación de espacios de Hilbert en la serie BVE-SP. Construcción propia. 61

LISTA DE TABLAS

Tabla 1 Descripción tipo de datos base BVE 22
Tabla 2 Tipos de datos acción AMZN 24
Tabla 3 Tabla comparativa de resultados del MSE para la serie Apple 53
Tabla 4 Tabla comparativa de resultados del MSE para la serie AMZN 55
Tabla 5 Tabla comparativa de resultados del MSE para la serie BVE-SP 56
Tabla 6 Resultados MSE para el modelo AR con EWMA original y en espacios de Hilbert para la serie APPL 57
Tabla 7 Resultados MSE para el modelo AR con EWMA original y en espacios de Hilbert para la serie AMZN 59
Tabla 8 Resultados MSE para el modelo AR con EWMA original y en espacios de Hilbert para la serie BVE-SP 60
Tabla 9 Resultados evaluación del modelo para la serie APPL 64
Tabla 10 Resultados evaluación del modelo para la serie AMZN 65
Tabla 11 Resultados evaluación del modelo para la serie BVE-SP 66

INTRODUCCIÓN

En el dinámico entorno financiero actual, la capacidad de anticipar y comprender la volatilidad de los mercados es esencial para la toma de decisiones informadas y la gestión de riesgos. En este contexto, surge la necesidad imperante de desarrollar metodologías innovadoras que no sólo aborden los desafíos inherentes a la complejidad de las series temporales financieras, sino que también proporcionen herramientas analíticas avanzadas para una estimación precisa de la volatilidad[1].

La volatilidad, como medida de la variabilidad de los precios de los activos, juega un papel crucial en la toma de decisiones de inversión y en la gestión de riesgos. Los modelos tradicionales, como ARCH y GARCH, han sido ampliamente utilizados para capturar la heterocedasticidad condicional de las series de tiempo financieras [2]. Sin embargo, estas metodologías presentan ciertas limitaciones, como la dificultad para modelar estructuras de dependencia más complejas y la suposición de que los retornos siguen una distribución normal [3].

La importancia de este proyecto radica en la necesidad crítica de superar las limitaciones de las metodologías convencionales, que a menudo subestiman la complejidad inherente de los datos financieros. La volatilidad, siendo una medida clave de riesgo, impacta, por ejemplo, directamente en las estrategias de inversión y la estabilidad del mercado; por ende, la capacidad de desarrollar modelos avanzados que capturen de manera precisa los patrones de volatilidad tiene implicaciones significativas en la toma de decisiones financieras y la gestión de carteras[4]. En este escenario, la investigación y aplicación de modelos autorregresivos en espacios de Hilbert con el uso de kernels reproductivos representan un paso hacia adelante en la búsqueda de enfoques más sofisticados y precisos. La teoría de espacios de Hilbert ofrece un marco matemático sólido, mientras que la aplicación de kernels reproductivos mejora la representación del modelo, permitiendo una estimación más eficaz de la volatilidad en contextos financieros complejos[5].

Para alcanzar este objetivo, se llevó a cabo un análisis exhaustivo de tres series de tiempo financieras, que incluyó el preprocesamiento, la selección de características relevantes y la construcción del modelo, realizando la comparación en la predicción de la volatilidad con un modelo autorregresivo (AR), modelo AR utilizando suavizamiento EWMA, y el modelo ARCH; sobre los cuales posteriormente se realizó la aplicación de kernels reproductivos. Además, se realizó una evaluación rigurosa del desempeño del modelo utilizando el Error Cuadrático Medio (MSE) el Error Absoluto Medio (MAE), el Coeficiente de Determinación (R^2) y una validación cruzada sobre el MSE.

I. DEFINICIÓN DEL PROBLEMA

El presente proyecto se enfoca en la exploración y análisis de medidas de volatilidad en el contexto de modelos autorregresivos para series temporales, específicamente dentro del marco teórico de los espacios de Hilbert. Para trazar un camino o ruta de desarrollo a través de un modelo para la toma de decisiones en el ámbito financiero, esta exploración y análisis implican la consideración cuidadosa de cada paso, ya que dichas decisiones tienen consecuencias significativas en la dirección y resultado final. La incertidumbre inherente a este entorno financiero cotidiano motiva la investigación y aplicación de medidas de volatilidad que reflejen de manera precisa la dinámica de las series temporales autorregresivas, contribuyendo así al conocimiento y comprensión más profunda de la variabilidad en los mercados financieros. En términos simples, es como seguir la vida financiera de una acción, viendo cómo cambian sus valores a lo largo del tiempo [1]. Ahora, tenga en cuenta que una serie de tiempo es la guía vital para la toma de decisiones en finanzas, pero a su vez la capacidad para prever cómo el comportamiento de estas series se ha convertido en una misión nada sencilla y más que esto se ha vuelto un desafío apasionante con diversos componentes aprovechables en la predicción como lo son la volatilidad y el riesgo de mercado.

En el contexto de la volatilidad y el riesgo del mercado, la predicción de series de tiempo financieras es aún más importante ya que la volatilidad es una medida de la variabilidad de los precios de las acciones y el riesgo del mercado es la probabilidad de que una inversión pierda valor [2]. Tratando el tema con la visión de la estadística tradicional, históricamente tenemos los modelos ARCH o GARCH los cuales permitieron conocer características de la predicción de la volatilidad en series de tiempo tal como lo son el comportamiento no estacionario de las series[6], los cuales pueden ser base de la automatización del proceso de predicción junto con el uso de la ciencia de datos.

A. PLANTEAMIENTO DEL PROBLEMA

La volatilidad financiera es un fenómeno complejo que puede ser modelado mediante series temporales y modelos estadísticos tradicionales; no obstante, estos enfoques convencionales pueden ser optimizados con modelos autorregresivos. Estos son instrumentos confiables en el campo de las proyecciones ya que emplean valores pasados para proyectar valores venideros. A pesar de su eficacia, estos modelos se topan con retos significativos al tratar de prever series de tiempo que no siguen patrones estables o que tienen tendencia a cambios bruscos [7].

Con el objetivo de obtener resultados más exactos que se puedan implementar en el mercado, se han creado y explorado modelos más versátiles para lograr mitigar los cambios abruptos en las series de tiempo financieras empleadas por la población de inversionistas y activistas bursátiles. A pesar de ser prometedores, estos modelos aún requieren de un estudio exhaustivo para su uso eficaz en la predicción de series de tiempo [8].

El desarrollo o adaptación de un modelo que sea capaz de anticipar series de tiempo financieras con características no estacionarias y elevada volatilidad podría representar un recurso inestimable en el ámbito financiero que daría apertura a posibles mejoras en la toma de decisión; por ejemplo, al tener un modelo que provea datos robustos del comportamiento de una serie financiera podría brindar a los inversores una vista óptima para tomar decisiones más fundamentadas y minimizar el peligro de pérdidas en sus inversiones.

Actualmente, el reto principal radica en predecir la volatilidad y el riesgo de mercado con una proximidad adecuada a los datos en tiempo real. Aunque existen software como Risk Simulator, MatLab y páginas web como XLSTAT que permiten la predicción financiera, la actualización manual de estos datos puede causar retrasos y afectar la precisión de las predicciones [9]. El avance en la ciencia de datos ofrece nuevas perspectivas para abordar problemas transversales como la predicción de la volatilidad en series de tiempo financieras. Este proyecto busca desarrollar herramientas que permitan implementar un modelo capaz de mitigar la fluctuación y el riesgo de mercado de manera efectiva [10].

B. FORMULACIÓN DEL PROBLEMA

Con base en el planteamiento del problema anterior, es pertinente formular la siguiente pregunta de investigación: ¿Cómo estimar la volatilidad y riesgo de mercado de una serie financiera utilizando modelos autorregresivos y espacios de Hilbert, implementando herramientas basadas en ciencia de datos, que su vez sirvan como insumo a inversores para la toma de decisiones informadas en el ámbito financiero?

Además, debemos también cuestionarnos lo siguiente: ¿Cómo validar que los datos financieros de distintas fuentes son confiables para el desarrollo del proyecto teniendo en cuenta su tratamiento? ¿Cómo los modelos autorregresivos pueden ser optimizados mediante la aplicación de espacios de Hilbert para una mejor predicción de la volatilidad en una serie de tiempo financiera? ¿Cuáles son las medidas de desempeño ideales que permitan evaluar la confiabilidad del modelo desarrollado?

II. OBJETIVOS DEL PROYECTO

A. OBJETIVO GENERAL

Explorar un modelo para la estimación y predicción de la volatilidad utilizando procesos autorregresivos y embebimientos en espacios de Hilbert con kernel reproductivo.

B. OBJETIVOS ESPECÍFICOS

1. Desarrollar una metodología para el preprocesamiento y acondicionamiento de bases de datos de series temporales financieras, utilizando herramientas de analítica y visualización.
2. Construir un esquema de estimación de la volatilidad financiera a partir de la representación de modelos autorregresivos en espacios de Hilbert con kernel reproductivo.
3. Implementar una metodología de validación exhaustiva del modelo desarrollado utilizando bases de datos de series financieras disponibles, teniendo en cuenta medidas de desempeño comunes en el estado del arte.

III. MARCO TEÓRICO Y ANTECEDENTES

La comprensión profunda de los procesos que subyacen en las series de tiempo y, en particular, en la predicción de la volatilidad, requiere un conocimiento sólido en tres áreas fundamentales: modelos autorregresivos, teoría de la probabilidad y espacios de Hilbert. Los modelos autorregresivos, al basarse en la relación entre observaciones pasadas y presentes, constituyen una herramienta esencial para capturar patrones temporales en datos secuenciales. La teoría de la probabilidad, por otro lado, proporciona el marco matemático necesario para cuantificar la incertidumbre y modelar la variabilidad inherente a las series de tiempo. En este contexto, los espacios de Hilbert emergen como una herramienta poderosa que va más allá de las limitaciones de los enfoques tradicionales, permitiendo una representación más precisa de las complejas relaciones subyacentes en los datos temporales. Este marco teórico explora en detalle la intersección de estos conceptos, destacando su aplicación conjunta en la predicción de la volatilidad en series de tiempo, con énfasis en cómo la integración de modelos autorregresivos en espacios de Hilbert puede ofrecer mejoras significativas en la capacidad predictiva y la comprensión de los fenómenos volátiles.

A. MARCO TEÓRICO

1. SERIES DE TIEMPO FINANCIERAS

Las series de tiempo financieras son una secuencia de datos que representa los precios, rendimientos o volúmenes de un activo financiero a lo largo del tiempo. Estas se utilizan para analizar el comportamiento de los precios de los activos, identificar tendencias y patrones, y tomar decisiones de inversión. Las series de tiempo financieras pueden ser estacionarias o no estacionarias. Una serie de tiempo estacionaria es una serie de tiempo en la que las características estadísticas, como la media, la varianza y la autocorrelación, no cambian con el tiempo. Una serie de tiempo no estacionaria es una serie de tiempo en la que las características estadísticas cambian con el tiempo [1].

Las series de tiempo financieras no estacionarias deben transformarse para que sean estacionarias antes de que se puedan analizar. Hay varios tipos de transformaciones que se pueden utilizar para transformar series de tiempo no estacionarias, como la diferenciación, la transformación de Box-Cox y la transformación de Fourier. Una vez que una serie de tiempo financiera se ha transformado para que sea estacionaria, se puede analizar utilizando una variedad de técnicas. Estas técnicas pueden utilizarse para identificar tendencias y patrones en los datos, y para predecir el comportamiento futuro de los precios [4].

2. VOLATILIDAD

La volatilidad se define como la variabilidad de los precios de un activo financiero. Se puede medir de varias maneras, pero la forma más común es utilizar la desviación estándar de los rendimientos diarios de un activo. La desviación estándar es una medida de la dispersión de los datos alrededor de la media. En el contexto financiero, la volatilidad se expresa generalmente como la desviación estándar de los rendimientos diarios de un activo, calculados a partir de una serie de tiempo de precios [4].

La volatilidad se puede medir de diversas maneras, siendo la desviación estándar una de las métricas más comunes. También se utilizan enfoques más avanzados, como la volatilidad implícita derivada de opciones financieras o la volatilidad histórica, que evalúa la variabilidad en función de datos pasados [5].

3. MODELOS AUTORREGRESIVOS

Los modelos autorregresivos (AR) son un tipo de modelo estadístico que se utiliza para predecir valores futuros de una variable a partir de sus valores pasados. Un modelo AR de orden p se puede expresar de la siguiente manera [4]:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \cdots + \beta_n y_{t-n} + \varepsilon_t \quad (1)$$

Donde:

y_t es el valor de la variable en el momento t.
β_0 es la constante.
$\beta_1, \beta_2, \ldots, \beta_n$ son los coeficientes de los términos autorregresivos.
ε_t es el término de error.

Los modelos AR como se muestra en la ecuación (1), se basan en la idea de que los valores futuros de una variable están relacionados con sus valores pasados. Por ejemplo, si los precios de las acciones suben hoy, es probable que sigan subiendo mañana. Los modelos AR capturan esta relación utilizando los valores pasados de la variable para predecir sus valores futuros.

Los modelos AR se pueden utilizar para predecir la volatilidad de varias maneras, de la cual una forma es utilizar un modelo AR para predecir los rendimientos de un activo. La volatilidad se puede calcular a partir de los rendimientos. Otra forma es utilizar un modelo AR para predecir directamente la volatilidad; pero, estos modelos pueden ser sensibles a los valores iniciales de la serie de tiempo y pueden tener problemas para capturar tendencias a largo plazo en la volatilidad [6].

4. MODELO AR

Los modelos autorregresivos (AR) son herramientas estadísticas ampliamente utilizadas para predecir valores futuros en series de tiempo financieras. Estos modelos se basan en la premisa de que los valores pasados influyen en los valores actuales, lo que hace que el método estadístico sea popular para analizar la naturaleza, la economía y otros procesos que cambian con el tiempo[4].

La fórmula general (2) para un modelo AR (p) es:

$$y_t = \delta + \sum_{i=1}^{p} \phi_i y_{t-i} + A_t \quad (2)$$

Donde:

- y_{t-i} son los valores pasados de la serie (retrasos).
- A_t es ruido blanco (aleatoriedad).
- δ se define mediante la siguiente ecuación:

$$\delta = \mu(1 - \sum_{i=1}^{p} \phi_i) \quad (3)$$

- p es el orden del modelo (por ejemplo, AR(1) es un proceso autorregresivo de primer orden).

Los modelos autorregresivos (AR) ofrecen una herramienta sencilla e interpretable para predecir la volatilidad en series de tiempo financieras. Al capturar la dependencia de los valores actuales en los pasados, estos modelos permiten identificar patrones y tendencias en la volatilidad [7]. Sin embargo, los modelos AR presentan ciertas limitantes ya que asumen una varianza constante en el tiempo, lo cual es poco realista en los mercados financieros [2]. Además, no pueden capturar efectos como el apalancamiento, donde la volatilidad tiende a aumentar en períodos de baja rentabilidad.

Para abordar estas limitaciones, se han desarrollado extensiones de los modelos AR, como los modelos GARCH (Generalized Autoregressive Conditional Heteroskedasticity), que permiten la modelización de una varianza que cambia con el tiempo [34]. Estos modelos han demostrado ser más efectivos para capturar la volatilidad en los mercados financieros, ya que consideran la heteroscedasticidad, es decir, la variabilidad de la volatilidad a lo largo del tiempo [2]. Además, los modelos GARCH pueden capturar fenómenos como el efecto apalancamiento, que es crucial para una representación más precisa de las series temporales financieras [35].

5. SUAVIZAMIENTO CON MEDIA MÓVIL EXPONENCIALMENTE PONDERADA (EWMA)

La media móvil exponencialmente ponderado (EWMA) es una técnica estadística utilizada para suavizar series de tiempo y resaltar las tendencias subyacentes. A diferencia de la media móvil simple, que asigna el mismo peso a todos los datos dentro de una ventana de tiempo, la EWMA otorga mayor importancia a los datos más recientes. El modelo EWMA se puede representar[4]:

$$\sigma_t^2 = (1-\lambda)\sum_{j=0}^{\infty}\lambda^j\left(r_{t-j}-r^-\right)^2 \quad (4)$$

Donde:

σ_t^2 es la estimación de la varianza para el periodo t
r^- es el rendimiento promedio de las observaciones
λ es el "factor de desintegración" que determina el peso de las observaciones recientes contra las antiguas.

Los modelos EWMA (Average Moving Exponential Weighted) proporcionan una serie de beneficios para el estudio de series temporales. Su mayor virtud reside en su habilidad para ajustarse con rapidez a las variaciones en la información, debido a la ponderación exponencial que concede mayor relevancia a las observaciones más recientes [36]. Esto los convierte en perfectos para seguir tendencias a corto plazo y identificar irregularidades. Además, su implementación y comprensión son relativamente fáciles, lo que los hace una herramienta común en varios ámbitos como finanzas, economía y control de calidad [37].

Aunque los modelos EWMA ofrecen algunas ventajas, también tienen ciertas restricciones. Un inconveniente clave es su susceptibilidad al parámetro de suavización (α). El seleccionar un valor incorrecto para α puede conducir a resultados sesgados o a una disminución de la información [36]. Adicionalmente, los modelos EWMA pueden resultar menos exactos que otras técnicas más sofisticadas, como los modelos ARIMA, en circunstancias donde la serie temporal muestra patrones estacionales o cíclicos marcados. Finalmente, la EWMA asume que los datos se encuentran en estado estacionario, lo que podría no ser cierto en numerosas aplicaciones reales [6].

6. MODELO ARCH

Los modelos ARCH (Autorregressive Conditional Heteroscedasticity) son una clase de modelos estadísticos diseñados para modelar la volatilidad de series de tiempo financieras. A diferencia de los modelos tradicionales que asumen una varianza constante, los modelos ARCH reconocen que la volatilidad de los retornos financieros tiende a agruparse en el tiempo, es decir, los períodos de alta volatilidad suelen ser seguidos por otros períodos de alta volatilidad, y lo mismo ocurre con los períodos de baja volatilidad[3]. Estos modelos se representan de la siguiente manera:

$$\sigma^2{}_t = \omega + \alpha_1\varepsilon^2{}_{t-1} + \alpha_2\varepsilon^2{}_{t-2} + \ldots + \alpha q\varepsilon^2{}_{t-}q \quad (5)$$

Donde:

- σ^2_t: Varianza condicional en el tiempo t.
- ω: Constante no negativa.
- $\alpha_1, \alpha_2, ..., \alpha q$: Parámetros no negativos.
- ε_t: Shock en el tiempo t.

Los modelos ARCH son herramientas estadísticas ampliamente utilizadas para modelar la volatilidad en series de tiempo financieras. Estos modelos son especialmente útiles para capturar la heterocedasticidad condicional, es decir, la variación en la volatilidad a lo largo del tiempo, y el clustering de volatilidad, que se refiere a la tendencia de los períodos de alta y baja volatilidad a agruparse. Esta capacidad de adaptarse a cambios en la varianza los hace ideales para modelar fenómenos como los "racimos de volatilidad" comúnmente observados en los mercados financieros [2].

Sin embargo, el éxito en la aplicación de los modelos ARCH depende en gran medida de una correcta especificación y estimación. La elección del orden del modelo y de la distribución de los errores puede influir significativamente en la precisión de los pronósticos. Además, estos modelos pueden presentar limitaciones en la predicción a largo plazo y en la captura de efectos asimétricos como el apalancamiento, donde los shocks negativos tienen un mayor impacto en la volatilidad que los shocks positivos [38].

7. MODELO GARCH

Los modelos GARCH (Generalized AutoRegressive Conditional Heteroskedasticity) son una extensión de los modelos ARCH y constituyen una herramienta fundamental en el análisis de series de tiempo financieras. Estos modelos permiten capturar de manera más flexible y precisa la volatilidad condicional, es decir, la varianza de los retornos de una serie de tiempo en un momento dado, condicionado a la información disponible hasta ese momento. A diferencia de los modelos ARCH, los modelos GARCH permiten que la varianza condicional dependa de las varianzas condicionales pasadas, lo que los hace más adecuados para modelar patrones de volatilidad más complejos [4].

Una característica distintiva de los modelos GARCH es que la varianza condicional se expresa como una función lineal de los cuadrados de los shocks pasados y de las varianzas condicionales pasadas. Esto permite capturar la persistencia de la volatilidad, es decir, la tendencia de los períodos de alta volatilidad a ser seguidos por otros períodos de alta volatilidad[4]. Además, los modelos GARCH pueden incorporar efectos de apalancamiento, lo que significa que las malas noticias (shocks negativos) pueden tener un mayor impacto en la volatilidad futura que las buenas noticias (shocks positivos).

$$\sigma^2{}_t = \omega + \alpha_1 \varepsilon^2{}_{t-1} + \alpha_2 \varepsilon^2{}_{t-2} + \ldots + \alpha q \varepsilon^2{}_{t-}q + \beta_1 \sigma^2{}_{t-1} + \beta_2 \sigma^2{}_{t-2} + \ldots + \beta p \sigma^2{}_{t-}p \quad (6)$$

Donde:

- σ^2_t: Varianza condicional en el tiempo t.
- ω: Constante no negativa.
- $\alpha_1, \alpha_2, ..., \alpha q$: Parámetros no negativos asociados a los shocks pasados.
- $\beta_1, \beta_2, ..., \beta p$: Parámetros no negativos asociados a las varianzas condicionales pasadas.
- ε_t: Shock en el tiempo t.

Al modelar la varianza condicional como una función de sus valores pasados, los modelos GARCH permiten obtener pronósticos más precisos de la volatilidad futura, lo cual es fundamental para la gestión de riesgos y la toma de decisiones de inversión. Además, su flexibilidad permite adaptar el modelo a diferentes tipos de series de tiempo y a diversos patrones de volatilidad [2].

Sin embargo, los modelos GARCH también presentan algunas limitaciones. La elección de la especificación del modelo (orden del proceso ARMA, distribución de los errores) puede ser compleja y requiere un profundo conocimiento de los datos y de la teoría. Además, los modelos GARCH pueden ser sensibles a cambios estructurales en la serie de tiempo y pueden subestimar la volatilidad en períodos de alta volatilidad extrema [2].

8. ESPACIOS VECTORIALES

Un espacio vectorial es una estructura matemática que consiste en un conjunto de elementos, llamados vectores, junto con dos operaciones: la suma de vectores y la multiplicación de un vector por un escalar. Estos espacios cumplen con una serie de propiedades, como la cerradura bajo la suma y la multiplicación, la existencia de un vector cero y la existencia de inversos aditivos para cada vector. Los espacios vectoriales son esenciales en el análisis de series de tiempo y modelos autorregresivos, ya que proporcionan el marco matemático para representar y manipular secuencias de datos. En este contexto, las observaciones de una serie de tiempo pueden interpretarse como vectores en un espacio vectorial. Esto permite aplicar herramientas y conceptos del álgebra lineal, como la autocorrelación, de manera efectiva para modelar las relaciones entre las observaciones y facilitar la construcción de modelos autorregresivos [11].

9. ESPACIOS DE HILBERT

Los espacios de Hilbert son estructuras matemáticas que combinan conceptos de geometría y análisis funcional. Son espacios vectoriales equipados con un producto interno que satisface propiedades esenciales, como la completitud. La completitud implica que todas las sucesiones de Cauchy convergen dentro del espacio, proporcionando una base sólida para el análisis matemático y la teoría de funciones [12].

La integración de espacios de Hilbert en modelos autorregresivos aporta ventajas significativas. Al considerar funciones como elementos en un espacio de Hilbert, se puede aprovechar la riqueza de la teoría de Hilbert para representar relaciones más complejas entre las observaciones en series de tiempo. Esta perspectiva abre la puerta a la utilización de herramientas matemáticas avanzadas, como las transformadas de Fourier en espacios de Hilbert, para capturar patrones no lineales de manera más efectiva [12].

Los espacios de Hilbert proporcionan un marco teórico riguroso para el análisis de señales, como las series de tiempo financieras. Un espacio de Hilbert es un espacio vectorial con un producto interno que además es completo. Esto significa que toda sucesión de Cauchy en el espacio converge a un elemento del espacio [39]. El producto interno induce una norma y una noción de distancia, permitiendo medir similitudes entre elementos. La completitud de este espacio garantiza la existencia de soluciones a muchas ecuaciones diferenciales e integrales que surgen en aplicaciones [39].

Un espacio de Hilbert contiene las siguientes propiedades:

- Linealidad

$$\alpha x + \beta y, z = \alpha(x, z) + \beta(y, z) \tag{7}$$

Donde:

A. $x, y, z \in H$

B. $\alpha, \beta \in \mathbb{C}$

- Simetría

$$(x, y) = (y, x)^* \tag{8}$$

- Positividad definida

$$\begin{gathered} (x, x) \geq 0 \\ x \in H, y(x, x) = 0 \\ x = 0 \end{gathered} \tag{9}$$

- Completitud: Toda sucesión de Cauchy en H converge a un elemento en H.

10. TEOREMA DE MERCER

El teorema de Mercer es un resultado fundamental en el análisis funcional que establece una conexión profunda entre las funciones simétricas y kernels, y las expansiones en serie. Este teorema es de gran importancia en el aprendizaje automático, especialmente en el contexto de los métodos del núcleo, como las máquinas de vectores de soporte (SVM) [8]. De manera simplificada, el teorema de Mercer establece que cualquier función simétrica y definida positiva K(x, y) en un conjunto compacto X puede ser expresada como una serie convergente de productos de funciones [40]:

$$k(x,y) = \sum_{i=0}^{\infty} \lambda_i \phi_i(x) \phi_i(y) \tag{10}$$

Donde:

λ_i son valores propios no negativos
$\phi_i(x)$ son las funciones propias correspondientes que forman una base ortonormal en un espacio de Hilbert adecuado.

Este teorema afirma que un núcleo, que en esencia mide la similitud entre dos puntos x e y, puede descomponerse en una suma ponderada de productos de funciones. Cada término de la suma representa una "dimensión" en un espacio de características implícito, donde las funciones ϕ_i actúan como coordenadas [40].

Para la predicción de volatilidad en series de tiempo financieras, el Teorema de Mercer facilita la comprensión de las relaciones no lineales entre variables financieras. Al descomponer un kernel en sus funciones y valores propios, se pueden capturar patrones complejos y variaciones intrínsecas que no serían evidentes de otra manera [41].

Esta descomposición es particularmente útil cuando se aplican métodos de aprendizaje automático como los vectores de soporte (SVM), que dependen de la elección adecuada de un kernel. Gracias al Teorema de Mercer, es posible garantizar que los kernels utilizados sean definidos-positivos y que su descomposición espectral sea válida, lo cual mejora la robustez y la precisión de los modelos [40].

11. TEOREMA DE RIESZ

El teorema de Riesz es un resultado fundamental en el análisis funcional que establece una conexión profunda entre los espacios de Hilbert y sus duales. Existen varias versiones de este teorema, cada una adaptada a diferentes contextos, pero todas comparten la idea central de establecer una correspondencia biyectiva entre funcionales lineales continuos en un espacio de Hilbert y los elementos del mismo espacio [39].

$$F[f] = \langle g(x), f \rangle_H = f(x) \tag{11}$$

Donde todo funcional lineal continuo en un espacio de Hilbert puede ser representado como un producto interno con un vector fijo. Esto significa que el espacio dual de un espacio de Hilbert es isométricamente isomorfo al propio espacio [39].

Aplicado a la predicción de volatilidad en series de tiempo financieras, el Teorema de Riesz permite modelar y entender mejor las relaciones entre diferentes variables financieras. Al utilizar espacios de Hilbert, es posible representar de manera precisa los funcionales lineales que describen estas relaciones, lo que mejora la precisión y la robustez de los modelos autorregresivos [42].

La capacidad para manejar estas representaciones de forma eficiente es crucial en el contexto financiero, donde las series de datos suelen ser complejas y presentan variaciones significativas. El uso del Teorema de Riesz, por tanto, no solo proporciona una base teórica sólida sino que también contribuye a la mejora práctica de las predicciones financieras [43].

En el contexto de la predicción de volatilidad en series de tiempo financieras, el uso combinado del Teorema de Riesz y el Teorema de Mercer permite desarrollar modelos más precisos y robustos. La capacidad para representar funcionales lineales y descomponer kernels en sus componentes espectrales mejora significativamente la precisión y la robustez de los modelos, proporcionando una base teórica sólida para el análisis y la predicción de la volatilidad.

12. REDES NEURONALES

Las redes neuronales, inspiradas en el funcionamiento del cerebro humano, han emergido como una herramienta poderosa en el ámbito de la ciencia de datos y el aprendizaje automático. Estas estructuras computacionales, compuestas por capas de nodos interconectados, tienen la capacidad de aprender patrones complejos y realizar predicciones basadas en datos de entrada[13].

Su implementación en el análisis de series de tiempo, particularmente en espacios de Hilbert, se justifica por su capacidad inherente para capturar relaciones no lineales y modelar la complejidad subyacente en datos temporales. La representación de series de tiempo en espacios de Hilbert facilita la aplicación de técnicas matemáticas avanzadas, y las redes neuronales se destacan como una herramienta idónea para abordar la no linealidad y la alta dimensionalidad presentes en este contexto [13]. La flexibilidad de las redes neuronales las convierte en un instrumento valioso para la predicción y el análisis de tendencias en series temporales financieras, permitiendo una comprensión más profunda y precisa de la dinámica subyacente en estos conjuntos de datos complejos [14].

13. MODELOS DE BOSQUES ALEATORIOS

Los modelos de bosques aleatorios (RF) son una técnica de aprendizaje automático supervisado que se ha utilizado con éxito en una amplia gama de aplicaciones, incluida la predicción de series financieras. Los RF se basan en la construcción de un conjunto de árboles de decisión independientes, cada uno de estos se entrena con un subconjunto aleatorio de los datos de entrenamiento. La predicción de un nuevo dato se realiza mediante la combinación de las predicciones de los árboles individuales[15].

Los RF presentan una serie de ventajas sobre otros métodos de predicción de series financieras, como su capacidad para controlar el sobreajuste, su robustez a los datos ruidosos y su facilidad de implementación.
En particular, los RF se han utilizado con éxito para predecir una amplia gama de variables financieras, como precios de acciones, tipos de interés, volatilidad y flujos de caja.
Por ejemplo, en un estudio reciente, los RF se utilizaron para predecir los precios de las acciones de las empresas del índice S&P 500. Los resultados mostraron que los RF eran capaces de obtener predicciones más precisas que otros métodos de predicción, como los modelos ARIMA[15].

B. ANTECEDENTES

Para la construcción de los antecedentes se realizó un análisis bibliográfico sobre el resultado en la base bibliográfica SCOPUS para la búsqueda de proyectos sobre la predicción de la volatilidad en series de tiempo financieras. Sobre esta lista de proyectos encontrados, se realiza un análisis de palabras claves a lo largo del tiempo desde 2014 hasta 2022 como se presenta en la siguiente imagen, de lo que podemos concluir que la predicción de la volatilidad se ha venido transformando a lo largo del tiempo, iniciando con modelos GARCH, cadenas de Markov o simulaciones de Montecarlo (lado izquierdo del mapa); y en la actualidad teniendo enfoques en el uso de Deep Learning, Redes Neuronales y Machine Learning (lado derecho del mapa)

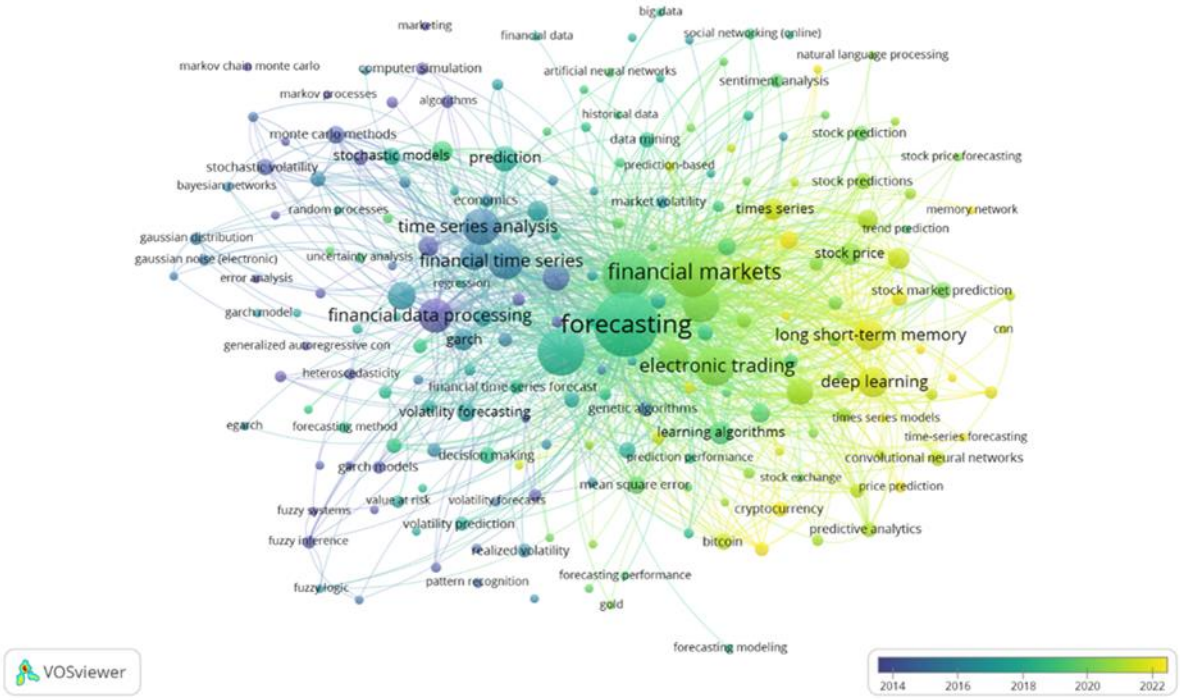

Figura 1 Co-ocurrencia de palabras clave en bibliografía relacionada 2014-2022. Construcción propia.

Partiendo de este análisis bibliométrico, el proyecto planteado toma un enfoque en la revisión de trabajos que se han desarrollado en esta misma área, un ejemplo de ello es el artículo presentado por Ortíz Rosas en 2022, este artículo que emplea modelos GARCH constituye un sólido estado del arte en la modelización de la volatilidad financiera. Sus hallazgos y metodologías sirven como punto de partida para el presente estudio, proporcionando un marco de referencia sólido para evaluar la eficacia de los modelos autorregresivos como lo son el Valor en Riesgo (VaR por sus siglas en inglés) o la Esperanza de Cola Condicional (TCE por sus siglas en inglés) [16]. Por otra parte, y en línea con el pronóstico de la volatilidad y el uso de medidas de riesgo, el trabajo de Vélez Hernández ofrece un valioso punto de referencia para el presente estudio. Su exploración de diversas combinaciones de modelos tradicionales y técnicas de aprendizaje automático sienta las bases para una investigación más profunda en este campo. Sin embargo, es fundamental considerar ciertas limitaciones como lo son la selección de los modelos híbridos y sus hiperparámetros puede ser un proceso altamente subjetivo y dependiente del conjunto de datos específico [17].

En otros artículos de estudios recientes sobre predicciones en el ámbito financiero Esteve y su artpicuko sobre la predicción de criptomonedas mediante inteligencia artificial representa un valioso punto de partida para nuestro estudio. Su enfoque en combinar modelos univariados y multivariantes, así como su análisis de pronósticos de densidad, proporciona una base sólida para explorar la predicción de volatilidad. Sin embargo, al centrarse en criptomonedas, el estudio presenta ciertas limitaciones al extrapolar sus

resultados a otros activos financieros. Además, la volatilidad, siendo un concepto más amplio que los retornos de los precios, requiere de modelos específicos que capturen su dinámica estocástica [18]. En cuanto al manejo de series no estacionarias, el artículo de Rodríguez y López [19]expone la metodología del suavizado exponencial es un método clásico y ampliamente utilizado en pronósticos. Sin embargo, es importante reconocer las limitaciones de este enfoque, el suavizado exponencial, si bien es sencillo de implementar, puede no capturar patrones más complejos presentes en la volatilidad financiera. Además, la elección de los parámetros de suavizamiento puede ser subjetiva y afectar significativamente los resultados.

En búsqueda de otros modelos que funcionen para la predicción en una serie de tiempo financiera, el artículo de Demirer y Gkillas sobre la predicción de la volatilidad del petróleo crudo utilizando bosques aleatorios empleando máquinas de aprendizaje para capturar patrones no lineales en los datos es innovador y ha demostrado ser eficaz en otros contextos. Además, su análisis de la aversión al riesgo como factor explicativo de la volatilidad aporta una dimensión adicional a nuestra investigación. Sin embargo, es importante considerar que su estudio se centra en un mercado específico (petróleo crudo) y que los bosques aleatorios, aunque versátiles, pueden ser sensibles a la sobre ajustamiento [20].

Sobre el uso de modelos con implementación de kernels reproductivos, el trabajo de López Pineda sobre modelos de regresión para datos funcionales en espacios de Hilbert con kernel reproductor aporta una base teórica sólida con su enfoque en la representación de datos funcionales en espacios de Hilbert [21]. Sin embargo, es importante destacar que el enfoque de López Pineda se centra en problemas de regresión, mientras que el objetivo de este proyecto es la predicción la volatilidad de una serie de tiempo financiera. Aunque la metodología de los kernels reproductores puede adaptarse a este contexto, es necesario realizar ajustes y consideraciones adicionales para abordar las particularidades de la predicción en series de tiempo. Por otra parte, el trabajo de Valencia Angulo sobre métricas entre procesos aleatorios en espacios de Hilbert con kernel reproductivo (RKHS) representa un avance significativo en la teoría de probabilidad y estadística. Su enfoque de embebimiento de distribuciones en RKHS ofrece una herramienta poderosa y flexible para comparar y clasificar procesos estocásticos [22]. Este artículo sienta las bases teóricas para el desarrollo de modelos autorregresivos en espacios de Hilbert, al proporcionar una forma rigurosa de medir distancias entre distribuciones de probabilidad asociadas a series de tiempo. Sin embargo, es importante notar que el trabajo de Valencia Angulo se enfoca en una teoría subyacente y no aborda directamente la predicción de volatilidad.

Adicional a lo anteriormente mencionado, Sang y Li ofrecen una opción en su artículo sobre regresión funcional no lineal empleando espacios de Hilbert con núcleo reproductivo (RKHS) [23]. No obstante, el alto costo computacional asociado a los métodos de RKHS puede ser un desafío para grandes conjuntos de datos, y la elección del núcleo adecuado puede requerir un análisis cuidadoso. Además, la aplicabilidad directa de este enfoque a series de tiempo financieras puede requerir adaptaciones, ya que las propiedades de las series financieras, como la heterocedasticidad y los efectos de contagio, pueden no ser capturadas completamente por modelos estáticos.

Es importante resaltar el trabajo realizado por por Villa Gil sobre la estimación y predicción del modelo ARCH para la volatilidad a través de embebimientos de distribuciones de probabilidad en espacios de Hilbert con kernel reproductivo [24]. Debido a que este antecedente se ajusta en gran forma al proyecto en gran medida a que el estudio explora la aplicación de técnicas de machine

learning a un modelo clásico de la econometría financiera, pero deja la puerta abierta a la exploración de modelos más sofisticados que capturen mejor las complejas dinámicas de la volatilidad, por lo tanto, con este proyecto al realizar previamente la comparación de diferentes modelos para la definición de la aplicación de kernels reproductivos se establece una vista más completa y compleja de la dinámica de kernenlización de acuerdo con la necesidad de cada modelo autorregresivo estudiado.

Continuo a esto es importante notar que, a lo largo de las últimas décadas, la modelización de series de tiempo financieras ha evolucionado significativamente, incorporando técnicas cada vez más sofisticadas para capturar la volatilidad. Entre los desarrollos más importantes se encuentra la implementación de espacios de Hilbert y análisis funcional, los cuales han permitido una comprensión más profunda y precisa de las dinámicas financieras.

Un estudio relevante en este ámbito es el de Bergmeir y Hyndman [25], quienes exploraron el uso de métodos de aprendizaje automático en la predicción de series temporales, destacando la importancia de seleccionar modelos adecuados que puedan manejar las características intrínsecas de los datos financieros. Sus investigaciones subrayan la necesidad de adaptar y personalizar los modelos para diferentes contextos financieros, lo que refuerza la relevancia de utilizar espacios de Hilbert y kernels en el presente estudio.

Además, Kuznetsov y Mohri [26] investigaron la aplicabilidad de los kernels en la regresión y clasificación de datos financieros, demostrando que los kernels no lineales pueden mejorar significativamente la precisión de las predicciones. Este trabajo es particularmente relevante para la presente tesis, ya que proporciona una base teórica y empírica sólida para la implementación de kernels reproductores en espacios de Hilbert.

Otro avance significativo es el trabajo de Hofmann, Schölkopf y Smola [27] sobre el aprendizaje con núcleos, donde se describe cómo los espacios de Hilbert con kernel reproductivo (RKHS) pueden ser utilizados para capturar relaciones complejas en datos financieros. Este enfoque ha sido aplicado exitosamente en diversos contextos, mostrando su eficacia en la predicción de volatilidad y otros indicadores financieros.

Por último, el artículo de Gilboa, Hutchinson y Poggio [28] sobre la implementación de métodos de aprendizaje profundo en series temporales financieras destaca la convergencia de técnicas de machine learning con métodos tradicionales de análisis financiero. Su investigación aporta una perspectiva novedosa al combinar redes neuronales con espacios de Hilbert, lo cual abre nuevas posibilidades para el presente estudio en términos de precisión y robustez en la predicción de volatilidad.

IV. METODOLOGÍA PARA EL PREPROCESAMIENTO DE UNA SERIE DE TIEMPO FINANCIERA

En este capítulo, la implementación de una metodología de preprocesamiento y acondicionamiento específicamente diseñado para bases de datos de series temporales financiera busca establecer un conjunto de prácticas y técnicas que permitan optimizar la calidad y relevancia de los datos financieros, preparándolos de manera efectiva para su posterior análisis en el marco del proyecto.

Se debe tener presente la definición de serie de tiempo financiera como secuencia de datos que representa los precios, rendimientos o volúmenes de un activo financiero a lo largo del tiempo[1] Por lo general, estas series de tiempo de manera gráfica pueden mostrar una tendencia a la baja, alza o plana, lo que permite comprender el movimiento del activo; todo esto enmarcado en una estacionalidad que se puede ver impactada por diferentes factores internos y externos al grupo económico al que pertenece la serie de tiempo.

A. EXPLORACIÓN DE FUENTES Y SERIES DE TIEMPO FINANCIERAS

Como punto de partida nos hemos centrado en el análisis y preprocesamiento de datos financieros de la Bolsa de Valores de Estambul (BVE) junto con otros siete índices internacionales clave: SP, DAX, FTSE, NIKKEI, BOVESPA, MSCE_EU y MSCI_EM, abarcando un período desde el 5 de junio de 2009 hasta el 22 de febrero de 2011 [29]. La elección de la Bolsa de Valores de Estambul como punto de partida para este análisis se justifica por su importancia como uno de los mercados financieros más prominentes en la región. Además, la disponibilidad de datos históricos detallados y la inclusión de índices internacionales importantes ofrecen una oportunidad única para desarrollar y validar metodologías de preprocesamiento y acondicionamiento de bases de datos de series temporales financieras [29].

La importancia de la Bolsa de Valores de Estambul como un mercado emergente clave en la región euroasiática, junto con la inclusión de otros índices internacionales prominentes, nos proporciona un panorama amplio y diverso de los movimientos del mercado durante el período de estudio. Este conjunto de datos ofrece una base sólida para desarrollar una metodología robusta de preprocesamiento y acondicionamiento que pueda ser aplicada posteriormente a otras bases de datos financieras similares, tanto en mercados emergentes como en mercados más establecidos. Ahora bien, para dar paso a los análisis preliminares del primer data set[29] es necesario un análisis descriptivo de los datos para que, de este modo, se pueda analizar el contexto y disponibilidad de información que sea de utilidad para poder proponer un primer modelo de ajuste y limpieza de la información en las series de tiempo a emplea; de tal manera que se obtuvo la siguiente Tabla 1 Descripción tipo de datos base BVE, que contiene el compilado de columnas y su descripción de los datos.

Tabla 1 Descripción tipo de datos base BVE

Nombre de la variable	**Role**	**Tipo**	**Valores faltantes**
Fecha	Característica	Fecha	No

ISE	Característica	Continuo	No
ISE	Característica	Continuo	No
SP	Característica	Binario	No
DAX	Característica	Continuo	No
FTSE	Característica	Continuo	No
NIKKEI	Característica	Continuo	No
BOVESPA	Característica	Continuo	No
UE	Característica	Continuo	No
EM	Característica	Continuo	No

Al realizar un análisis exploratorio sobre el set de datos se pueden derivar diversas inferencias significativas sobre el comportamiento de los índices bursátiles estudiados: DAX, FTSE, NIKKEI, BOVESPA, EU y EM ya mencionados anteriormente. Los valores medios de estos índices sugieren ligeras tendencias al alza, con promedios diarios de retornos positivos, aunque marginales, situándose todos ellos entre 0.000308 (NIKKEI) y 0.000936 (EM). Esto indica una tendencia general de crecimiento en los mercados, aunque las variaciones diarias son relativamente pequeñas. Los índices BOVESPA y EM presentan las mayores medias de retorno, lo cual podría señalar un mayor rendimiento en mercados emergentes comparado con los desarrollados.

En este contexto, la base de datos de la Bolsa de Valores de Estambul es una herramienta valiosa. Consta de 536 entradas, que incluyen varios índices financieros sin valores nulos, como ISE, SP, DAX, FTSE, NIKKEI, BOVESPA, EU y EM. Este conjunto de datos, distribuido en columnas de tipo datetime64[ns] y float64, con un uso de memoria aproximado de 42.0 KB, proporciona una base robusta y precisa para llevar a cabo análisis detallados y detectar tendencias significativas en los mercados financieros.

El análisis de la dispersión de los datos, a través de la desviación estándar y los cuartiles, revela importantes diferencias en la volatilidad de los índices. El índice NIKKEI muestra la mayor desviación estándar (0.014850), indicando una mayor variabilidad en sus retornos diarios en comparación con otros índices como el FTSE (0.012656) y el EU (0.012990). Adicionalmente, la comparación entre el rango intercuartílico y los valores extremos (mínimos y máximos) sugiere que, a pesar de estas fluctuaciones, la mayoría de los movimientos diarios de los índices permanecen dentro de un margen predecible, aunque existen eventos extremos que provocan mayores desviaciones. Estos hallazgos son esenciales para los inversores y analistas financieros, ya que subrayan la necesidad de considerar tanto las tendencias generales de crecimiento como la volatilidad específica de cada mercado en sus estrategias de inversión. Una muestra de lo anterior se refleja en la siguiente matriz de correlación en la Figura 2 Matriz de correlación BVE.

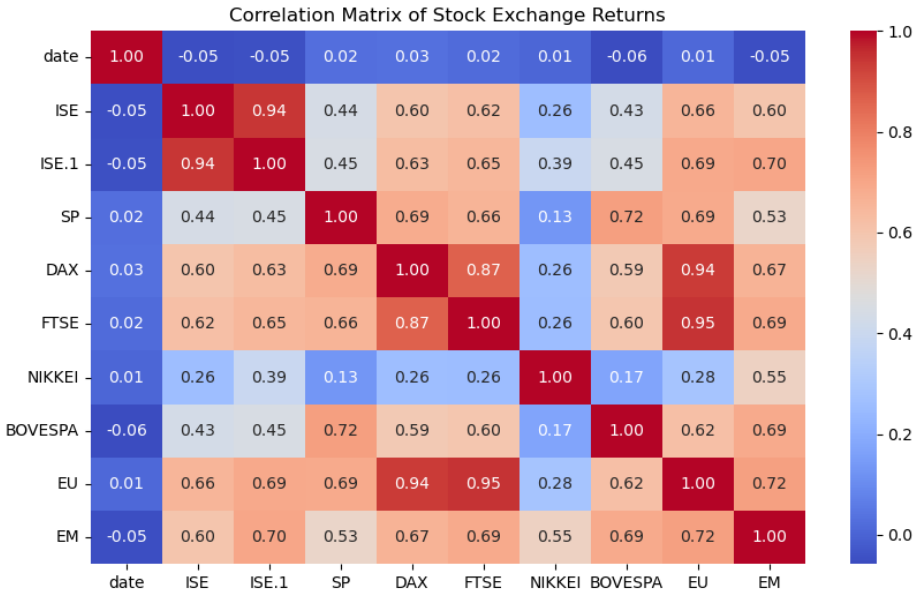

Figura 2 Matriz de correlación BVE. Construcción propia. Fuente: Serie de tiempo BVE

Adicional a esto, se logró evidenciar que, el índice bursátil de Estambul (BIST 100) históricamente ha mostrado una notable volatilidad, influenciada por factores tanto internos como externos. A nivel nacional, las fluctuaciones en la política económica y las decisiones del banco central juegan un papel crucial en la determinación del rendimiento del mercado. Además, eventos políticos y sociales también pueden impactar significativamente la confianza del inversor y, por ende, los retornos del índice.

Por otra parte, dentro de las fuentes planteadas en el anteproyecto, se determinó Kaggle como una fuente propicia para la obtención de datos de libre acceso con historia que permita la modelación de la predicción de la volatilidad, objetivo principal del proyecto. Dentro de esta fuente de información encontramos la base de datos *Stock Market Data (NASDAQ, NYSE, S&P500),* la cual almacena información de los diferentes activos en distintos formatos como se muestra en la Tabla 2**¡Error! No se encuentra el origen de la referencia.**, y que se negocian en los indicadores mencionados con un histórico desde 2020 [30].

Tabla 2 Tipos de datos acción AMZN

#	**Columna**	**Non-Null Count**	**Dtype**
0	Low	1194	Float64
1	Open	1194	Float64
2	Volumen	1194	Float64
3	High	1194	Float64
4	Close	1194	Float64

5	Adjusted Close	1194	Float64

La base de datos de Amazon (AMZN) se convierte en una herramienta valiosa. Esta base de datos consta de 6438 entradas que incluyen varios índices financieros sin valores nulos, como la fecha (tipo datetime64[ns]), los precios de apertura, cierre, máximos y mínimos (tipo float64), el volumen de transacciones (tipo int64) y el cierre ajustado (tipo float64). Con un uso de memoria aproximado de 352.2 KB, este conjunto de datos proporciona una base robusta y precisa para llevar a cabo análisis detallados y detectar tendencias significativas en los mercados financieros.

Para la serie se realizó un análisis de la matriz de correlación como se muestra en la Figura 3, teniendo los siguientes resultados, donde sugieren que el precio de la acción AMZN está fuertemente correlacionado con las variables Bajo, Abierto, Alto y Cierre. Esto significa que cualquiera de estas variables puede ser útiles para predecir el precio futuro de la acción ya que representan de manera significativa el comportamiento del precio de la acción. La correlación negativa débil entre el precio de la acción y el Volumen sugiere que esta variable puede no ser tan útil para la predicción del precio futuro de la acción.

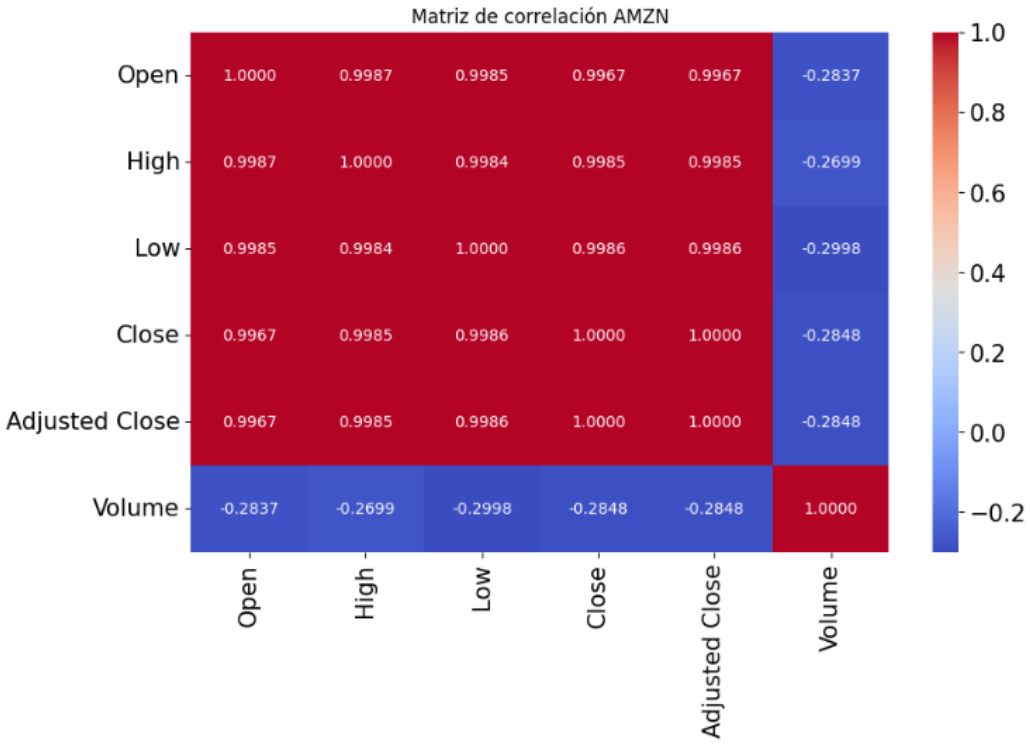

Figura 3 Matriz de correlación de la acción Amazon (AMZN). Construcción propia. Fuente: Serie de tiempo AMZN

Adicional a las series anteriores y para tener diferentes puntos de comparación posteriormente en el desarrollo de los modelos, se decidió incluir una tercera serie de tiempo, la cual contiene los precios de la acción Apple desde 2020, como se puede apreciar en la Figura 5, se presenta un histograma que muestra la distribución de los precios de cierre de las acciones de Apple. Cada barra representa un rango de precios y su altura indica la frecuencia con la que el precio de cierre cayó dentro de ese rango. Además, se

ha superpuesto una curva de densidad para visualizar de manera más suave la distribución de los datos. La distribución parece ser aproximadamente normal, con una ligera asimetría hacia la derecha. Esto sugiere que la mayoría de los precios de cierre se concentran alrededor de un valor central, pero hay algunos precios más altos que ocurren con menor frecuencia. La amplitud de la distribución, o la dispersión, sugiere que los precios de cierre han fluctuado en un rango considerable. La curva de densidad se extiende desde aproximadamente 50 dólares hasta 225 dólares, lo que indica una variabilidad significativa en los precios siendo un insumo adecuado para la estimación de la volatilidad con distintos modelos que se plantean posteriormente.

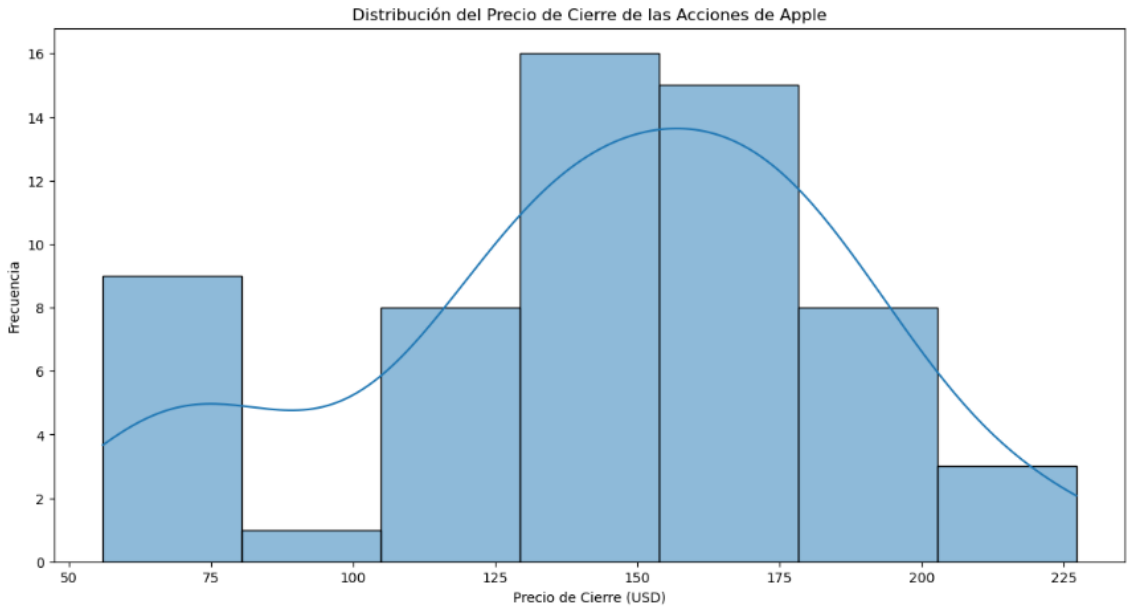

Figura 4 Distribución de precios sobre la acción APPL. Construcción propia. Fuente: Serie de tiempo APPL.

El conjunto de datos de Apple (AAPL), con sus 1200 entradas, es manejable y eficiente en términos de uso de memoria, ocupando aproximadamente 75.0 KB. Contiene información sobre fechas, precios de apertura y cierre, máximos, mínimos y volumen de transacciones. No presenta valores faltantes, es decir cumple con lo necesario para seruna base de datos funcional apta como herramienta valiosa para realizar análisis básicos y estudiar tendencias a lo largo del tiempo.

B. METODOLOGÍA PARA EL PREPROCESAMIENTO DE SERIES DE TIEMPO FINANCIERAS

Con el objetivo de preparar los datos para un posterior análisis, se describen los pasos más comunes para el preprocesamiento de series de tiempo financieras:

1. LIMPIEZA DE DATOS

Este método se compone de tres puntos para obtener una base de datos limpia para su uso [31]:

1) Detección y tratamiento de valores atípicos: Estos datos, que se encuentran significativamente alejados del resto de los valores, pueden distorsionar los resultados del análisis. Técnicas como el Z-score o el rango intercuartílico permiten identificar y tratar estos valores atípicos mediante su eliminación o sustitución por valores más representativos [31].
2) Tratamiento de datos faltantes que se pueden presentar por diversas razones (errores de entrada de datos, problemas de transmisión, pérdida de información), sobre los cuáles se pueden implementar métodos de imputación o la eliminación de registros, dependiendo de un previo análisis de la base de datos, ya que la ausencia de datos puede introducir sesgos en el análisis. Para abordar este problema, se pueden emplear diversas técnicas de imputación, como la media, la mediana, la moda, la interpolación lineal o métodos más sofisticados basados en modelos predictivos [31].
3) Corrección de errores al momento de la medición que pueden introducir ruido a la base de datos, los cuales pueden partir desde la captura o el procesamiento de los datos y se puede tratar con técnicas como la validación cruzada y la comparación con fuentes de datos externas [31].

2. TRANSFORMACIÓN DE LOS DATOS

Una base de datos puede ser transformada en el momento de su limpieza y posteriormente de acuerdo con las necesidades y la elección del administrador. Algunas de las formas de transformación de los datos pueden ser[31]:

1) La normalización y estandarización se utilizan para escalar los datos a un rango específico, como entre 0 y 1 o entre -1 y 1. Esto puede ser útil para mejorar la comparabilidad de las variables y el rendimiento de los modelos de aprendizaje automático.
2) La transformación logarítmica se utiliza para transformar datos que tienen una distribución sesgada hacia la derecha. Esto puede ayudar a normalizar la distribución de los datos y estabilizar la varianza.
3) La diferenciación se utiliza para eliminar la tendencia de una serie de tiempo. Esto puede ser útil para analizar los cambios en la tasa de crecimiento de la serie.

3. MANEJO DE LA ESTACIONALIDAD

Como se mencionó al inicio de este capítulo, una de las características de las series de tiempo financieras es la estacionalidad de los datos. Para algunos casos es necesario aplicar un manejo a esta característica que puede ser[31]:

1) Utilizar la desestacionalización para eliminar la estacionalidad de una serie de tiempo.
2) El modelado de la estacionalidad implica incluir términos estacionales en un modelo de pronóstico. Esto puede ser útil para capturar los patrones estacionales de la serie de tiempo.

4. VALIDACIÓN DE DATOS

Finalmente, para un correcto preprocesamiento de una base de datos sobre una serie de tiempo financiera, es necesario tener

presente el análisis de gráficos que permite identificar patrones y otros problemas que hayan pasado por alto en los pasos anteriores; y la realización de estadísticos de prueba para evaluar la normalidad o la correlación de los datos, por ejemplo.

Aplicando este paso a los datos seleccionados, se obtienen los siguientes resultados para la información de la bolsa de Estambul donde externamente, la economía global y las condiciones del mercado emergente afectan considerablemente el BIST 100. Similar a otros índices de mercados emergentes, el BIST 100 tiende a ofrecer rendimientos potencialmente más altos, aunque con un nivel de riesgo igualmente elevado. Esto se alinea con la tendencia observada en los rendimientos medios positivos, pero con alta volatilidad en los índices emergentes como BOVESPA y EM en el análisis anterior, una muestra a ello se refleja a continuación en la Figura 6.

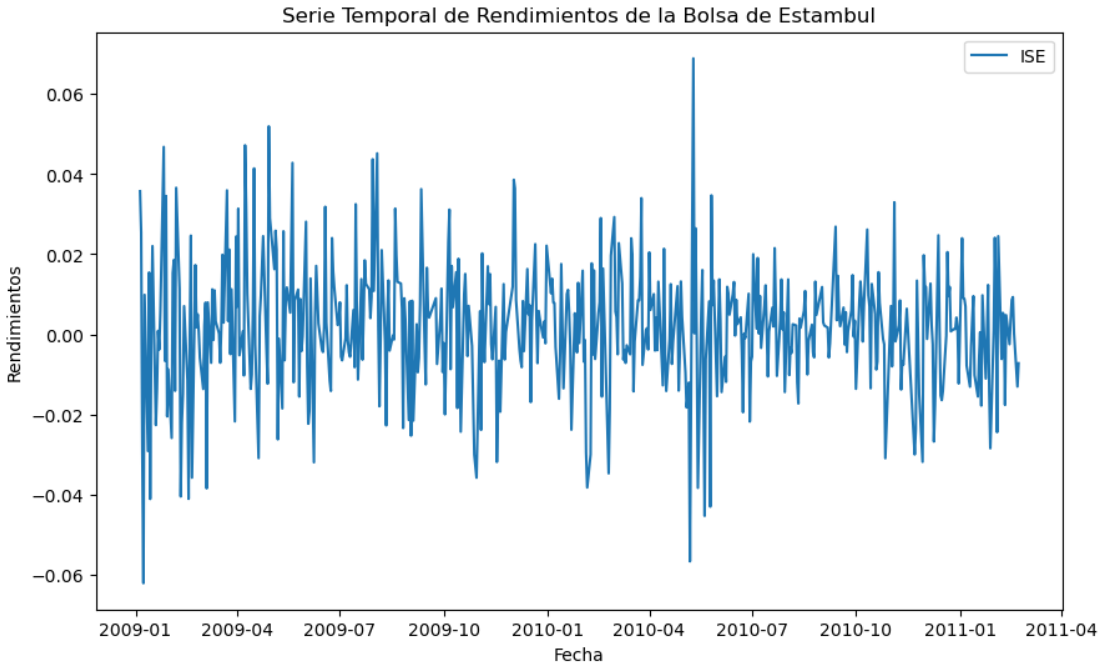

Figura 5 Serie de tiempo de rendimientos de la Bolsa de Estambul. Construcción propia. Fuente: Serie de tiempo BVE.

Los índices bursátiles, como el NASDAQ, NYSE y S&P 500, son indicadores económicos cruciales que reflejan el desempeño general del mercado de valores. Estos índices no solo sirven como puntos de referencia para los inversores, sino que también proporcionan información valiosa para la predicción de la volatilidad de acciones individuales y del mercado en general. A su vez estos índices a largo plazo pueden reflejar el crecimiento económico general, las condiciones de la industria y otros factores macroeconómicos. La volatilidad de las acciones individuales por índice no solo está influenciada por factores específicos de la empresa, sino también por factores sistémicos que afectan a todo el mercado, por lo tanto, al basar el proyecto en un índice se tiene una amplia gama de posibilidades y mercados.

Por otra parte, la acción AMZN ha tenido un comportamiento creciente a lo largo del tiempo, peor con un cambio en la tendencia

desde el impacto del Covid-19 en las operaciones de la empresa como se puede apreciar en la Figura 7 a continuación:

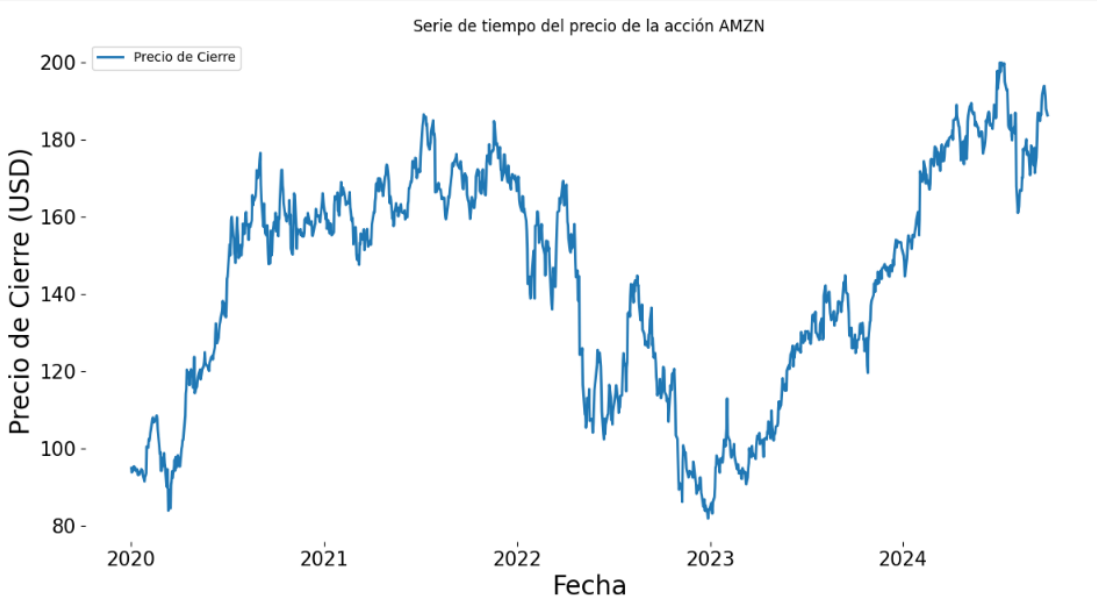

Figura 6 Serie de tiempo de precios de la acción Amazon (AMZN). Construcción propia. Fuente: Serie de tiempo AMZN.

A pesar de la tendencia inicial, al realizar un análisis de retardo de la serie de tiempo donde el eje X representa el valor de la serie de tiempo en el año t, mientras que el eje Y representa el valor de la serie de tiempo en el año t + 1. La serie de tiempo de Amazon es una serie de tiempo no estacionaria con una tendencia positiva y una gran cantidad de variabilidad visible en el análisis de retardo de la Figura 8.

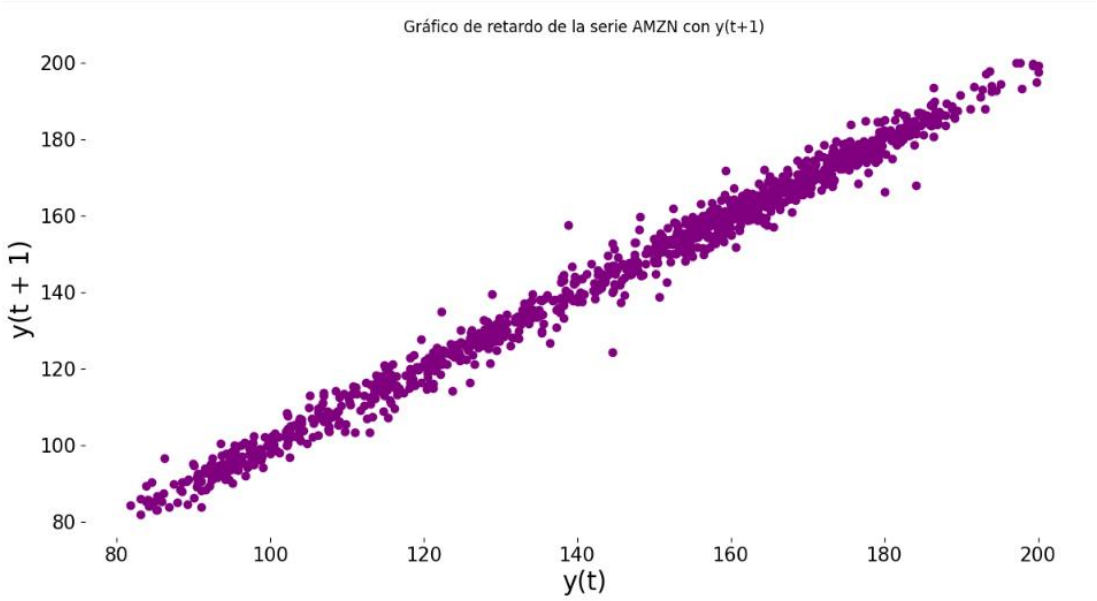

Figura 7 Análsis de retardo de la serie de tiempo Amazon (AMZN) Construcción propia. Fuente: Serie de tiempo AMZN.

Finalmente, hemos realizado un análisis preliminar de la gráfica del precio de las acciones de Apple desde 2020 hasta julio de 2024 como se muestra en la figura 9. Observamos una tendencia general al alza, lo que sugiere un buen desempeño de la compañía en el mercado. Sin embargo, también identificamos períodos de volatilidad y puntos de inflexión que podrían indicar cambios en la tendencia.

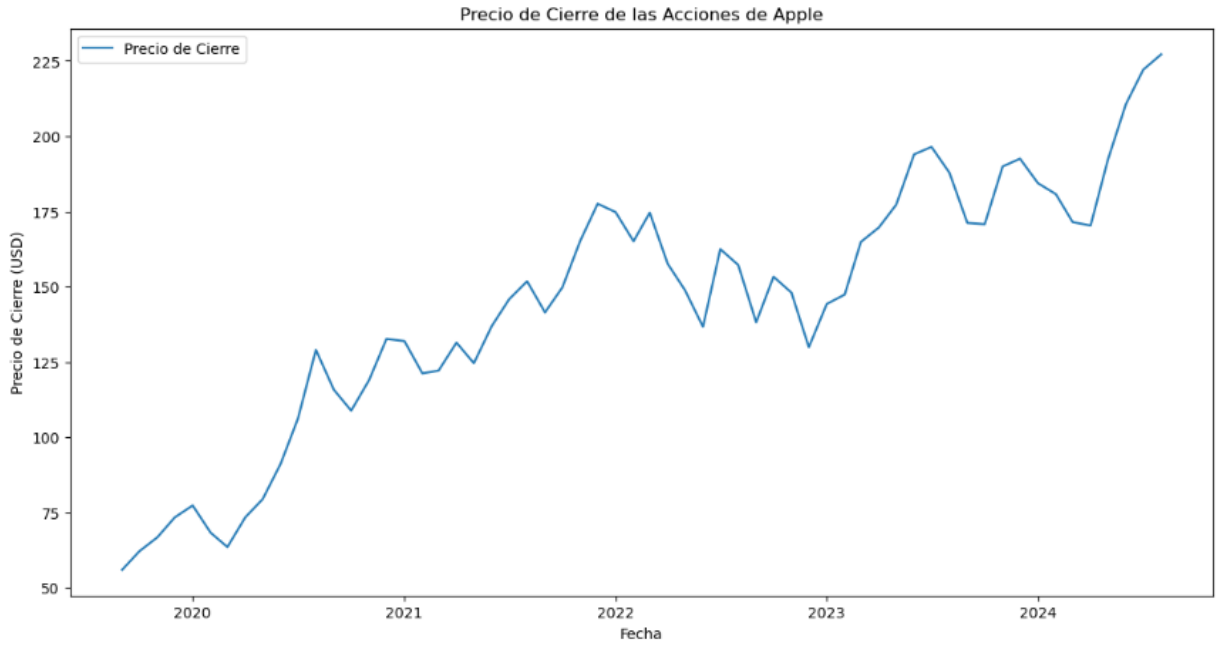

Figura 8 Precios de la acción APPL. Construcción propia. Fuente: Serie de tiempo APPL.

V. MODELOS AUTORREGRESIVOS Y ESPACIOS DE HILBERT

Los modelos autorregresivos (AR) son herramientas estadísticas ampliamente utilizadas para predecir valores futuros en series de tiempo financieras. Estos modelos se basan en la idea de que cada valor en una serie temporal depende linealmente de sus valores anteriores. En particular, los modelos AR consideran que el valor de la variable de resultado en un momento t está relacionado con los valores previos de la misma serie y una componente de error aleatorio[32]. Los modelos autorregresivos (AR) han sido durante décadas una herramienta fundamental en el análisis de series de tiempo, permitiendo modelar la evolución de una variable en función de sus valores pasados. Sin embargo, la capacidad de estos modelos para capturar patrones complejos y no lineales puede verse limitada.

Recientemente, la teoría de los espacios de Hilbert de funciones reproductivas (RKHS) ha emergido como una poderosa herramienta para el análisis de datos, ofreciendo una forma flexible y no paramétrica de modelar relaciones funcionales. La combinación de modelos AR con la teoría de RKHS ha dado lugar a una nueva clase de modelos que pueden capturar de manera más precisa y flexible las dinámicas de series de tiempo.

En el contexto de la predicción de volatilidad, los modelos AR pueden ser especialmente útiles. La volatilidad es una medida de la dispersión o variabilidad de los rendimientos financieros, y su predicción es crucial para la gestión de riesgos y la toma de decisiones en los mercados financieros[33]. Al aplicar modelos AR en espacios de Hilbert, se aprovecha la estructura matemática de estos espacios para modelar la relación entre los valores pasados y futuros de la volatilidad.

A. ESPACIO DE HILBERT CON KERNEL REPRODUCTIVO

Un RKHS (Reproducing kernel Hilbert Space en inglés), es un tipo especial de espacio de Hilbert de funciones que posee una propiedad particular: la capacidad de reproducir cualquier función del espacio a través de un producto interno con una función especial llamada núcleo reproductor. Sea H un espacio de Hilbert de funciones y k una función $X \rightarrow \mathbb{R}$, la función k es un kernel reproductivo y H un RKHS si [44]:

- $\forall x \in X, k_x \in H$. Es decir: $k_x: y \rightarrow k(x,y), k_x(y) = k(x,y)\ donde\ k_x = k(x, \cdot)$
- K tiene la propiedad reproductiva: $\forall x \in X\ y\ g\ en\ H, g(x) = \langle g(\cdot),\ k_x \rangle$

Los RKHS han ganado una gran relevancia en el campo del aprendizaje automático, en particular en métodos como las máquinas de vectores de soporte (SVM) y los procesos Gaussianos. Adicionalmente, la relación entre un Espacio de Hilbert con Núcleo Reproductor (RKHS) y su núcleo reproductor es fundamental. El núcleo define la estructura del espacio y determina las propiedades de las funciones que pueden ser representadas en él [45].

B. TIPOS DE KERNELS REPRODUCTIVOS

Al transformar los datos en un espacio de características de mayor dimensión, los kernels habilitan la construcción de modelos

más flexibles y poderosos. En este trabajo, exploraremos los diferentes tipos de kernels comúnmente utilizados, como el lineal, polinomial, RBF y sigmoidal, analizando sus características, ventajas y desventajas. Comprender las propiedades de cada kernel es esencial para seleccionar el más adecuado para un problema específico y optimizar el rendimiento de los modelos de aprendizaje automático.

1. Kernel lineal: Es el más sencillo de todos y se calcula simplemente como el producto punto entre dos vectores. Este kernel es útil cuando se asume que los datos son linealmente separables. Aunque su simplicidad puede limitar su capacidad para capturar relaciones no lineales, es un buen punto de partida para muchos problemas [46]. Matemáticamente se expresa como:

$$k(x, y) = (k \cdot y) \tag{12}$$

2. Kernel polinomial: generaliza el kernel lineal al considerar combinaciones polinomiales de los atributos y permite modelar relaciones no lineales más complejas que el kernel lineal [46]. Matemáticamente se representa como:

$$k(x, y) = (y(x, y) + c)^d \tag{13}$$

   Donde:

   y: Es un parámetro que controla la influencia del producto punto.

   c: Es un parámetro que controla el término independiente

   d: Es el grado del polinomio

3. Kernel RBF: Kernel Radial Basis Function o Kernel Gaussiano, es uno de los kernels más utilizados debido a su capacidad para capturar relaciones no lineales altamente complejas. Este kernel mapea los datos a un espacio de dimensión infinita, lo que permite encontrar separadores no lineales muy sofisticados [47]. Matemáticamente se define como:

$$k(x, y) = \exp\left(-y\left|\left|x - y\right|\right|^2\right) \tag{14}$$

   Donde:

   y: Es un parámetro que controla la anchura de la función gaussiana.

   $\left|\left|x - y\right|\right|^2$: Es la distancia euclidiana al cuadrado entre dos vectores x e y.

4. Kernel sigmoidal: Está inspirado en la función sigmoide utilizada en las redes neuronales artificiales. Aunque este kernel ha sido utilizado en algunas aplicaciones, su desempeño no es tan consistente como el de otros kernels, y su elección suele ser más empírica [47]. Matemáticamente se define como:

$$k(x, y) = \tanh(y(x, y) + c) \tag{14}$$

Donde:

y: Es un parámetro que controla la pendiente de la función tangente hiperbólica.

c: Es un parámetro que controla el punto de inflexión.

VI. APLICACIÓN DE MODELOS AUTORREGRESIVOS A SERIES DE TIEMPO FINANCIERAS

En el presente capítulo se exponen detalladamente los resultados obtenidos al aplicar modelos autorregresivos a las series de tiempo financieras que fueron analizadas en el capítulo 4. El objetivo principal de este estudio es modelar y estimar el comportamiento de la volatilidad utilizando los modelos autorregresivos propuestos en el capítulo anterior. Estos modelos, seleccionados por su capacidad para capturar dinámicas complejas en datos financieros, se evalúan para determinar su eficacia en predecir la volatilidad futura. Además, se analiza cómo estos pronósticos pueden integrarse posteriormente con la implementación de espacios de Hilbert, con el fin de mejorar la precisión y robustez de las predicciones. Mediante este enfoque, se espera proporcionar una herramienta más potente para la toma de decisiones en el ámbito financiero.

Para llevar a cabo este análisis la secuencia de los ejercicios de modelado y estimación del comportamiento de las series de tiempo financieras primero fue necesario cargar y pre procesar los datos históricos de las series, así al filtrar las observaciones más recientes es posible reflejar mejor las condiciones actuales del mercado, no obstante, para lograr una óptima condición en el escenario de cada una de las series de tiempo fue necesaria la normalización de los datos mediante escalas de mínimos y máximos lo cual es crucial para asegurar que los modelos trabajen con datos comparables, mitigando posibles sesgos debido a diferentes escalas. La aplicación de la técnica de suavizado exponencial (EWMA) permite ponderar más los datos recientes y suavizar las fluctuaciones temporales, mejorando así la capacidad del modelo para reaccionar a cambios recientes en los datos. Al lograr identificar el mejor de los escenarios es necesario definir y entrenar los modelos autorregresivos (AR), así como modelos ARCH y GARCH, tanto sobre la serie de tiempo normalizada como la suavizada, permitiendo una evaluación integral de cómo diferentes métodos de suavizado impactan la precisión de las predicciones; de este modo al tener los modelos implementados es posible realizar la evaluación del desempeño de cada modelo la cual se realiza mediante el cálculo del error cuadrático medio (MSE), y las visualizaciones gráficas que facilitan la comparación intuitiva entre los valores reales y estimados. Este enfoque estructurado permite identificar el modelo más efectivo para predecir la volatilidad de las series de tiempo financieras, proporcionando una base sólida para su integración futura con espacios de Hilbert, y así ofrecer una herramienta robusta para la toma de decisiones en el ámbito financiero [48].

A. SERIE APPLE (APPL)

Para la serie de tiempo de los precios de la acción Apple (APPL) se realizó el desarrollo de los modelos autorregresivos de la siguiente manera y obteniendo los siguientes resultados. Para el modelo autorregresivo (AR), ajustado a los datos históricos de las acciones de Apple ha mostrado un Error Cuadrático Medio (MSE) de 0.0157. Este valor indica que el modelo tiene una muy alta precisión en la predicción de los precios de cierre. La presencia de ruido y volatilidad en los datos financieros es un factor que contribuye a un posible error en las posibles predicciones del modelo AR; aquí se siguen de cerca los valores reales en el conjunto de prueba, lo que sugiere que el modelo es capaz de capturar las tendencias generales de la serie de tiempo. Sin embargo, es importante considerar que los modelos AR pueden tener limitaciones en la captura de cambios estructurales y eventos inesperados

en el mercado, para evitar el ruido en la data y poder mitigar la volatilidad se tomó como métrica de exposición los últimos dos periodos conocidos por la serie de tiempo.

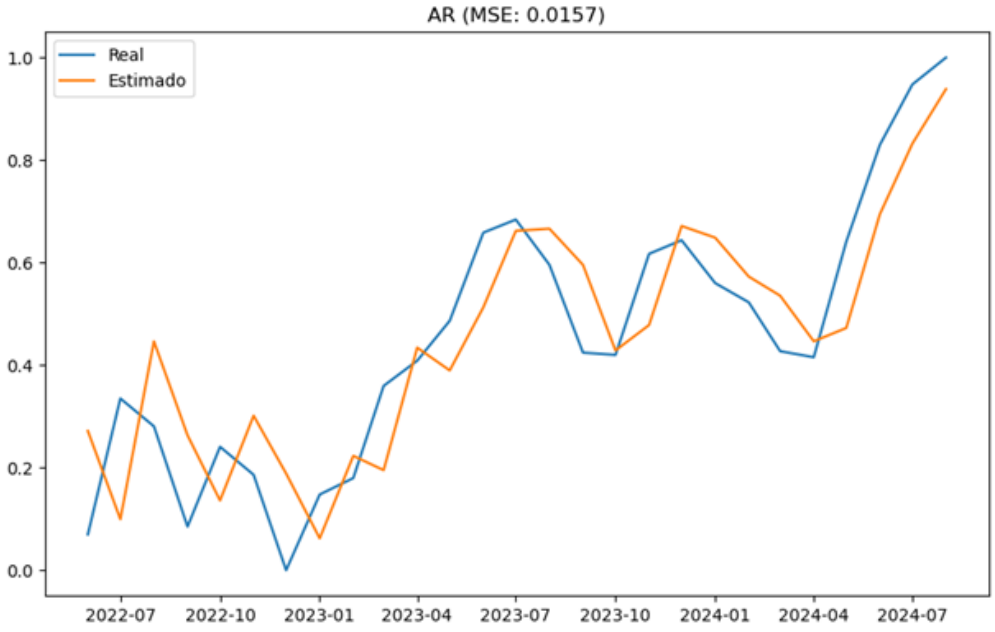

Figura 9 Resultados predicción de la volatilidad del precio de la acción APPL con modelo AR. Construcción propia.

Los gráficos de la función de autocorrelación (ACF) y la función de autocorrelación parcial (PACF) han mostrado patrones que confirman la elección del número de rezagos en el modelo AR. Estos gráficos son herramientas útiles para entender la estructura de la serie de tiempo y ajustar adecuadamente el modelo.

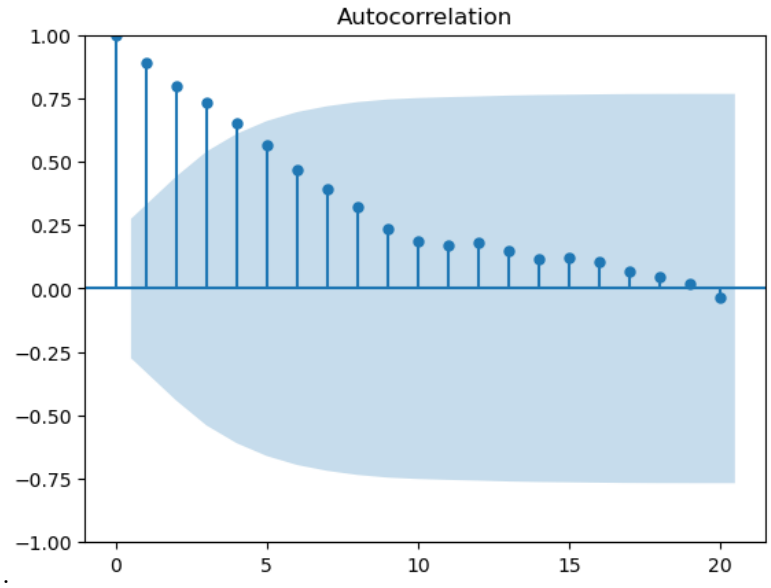

Figura 10 Gráfico de autocorrelación del modelo AR. Construcción propia.

El análisis de la serie de tiempo de las acciones de Apple (AAPL) utilizando un modelo de Media Móvil Exponencialmente Ponderada (EWMA) ha proporcionado una visión adicional sobre la dinámica de los precios de cierre. El modelo EWMA ha mostrado un Error Cuadrático Medio (MSE) de 0.0001. Este valor es significativamente más bajo que el MSE del modelo AR lo que indica que el modelo EWMA tiene una precisión superior en la predicción de los precios de cierre.

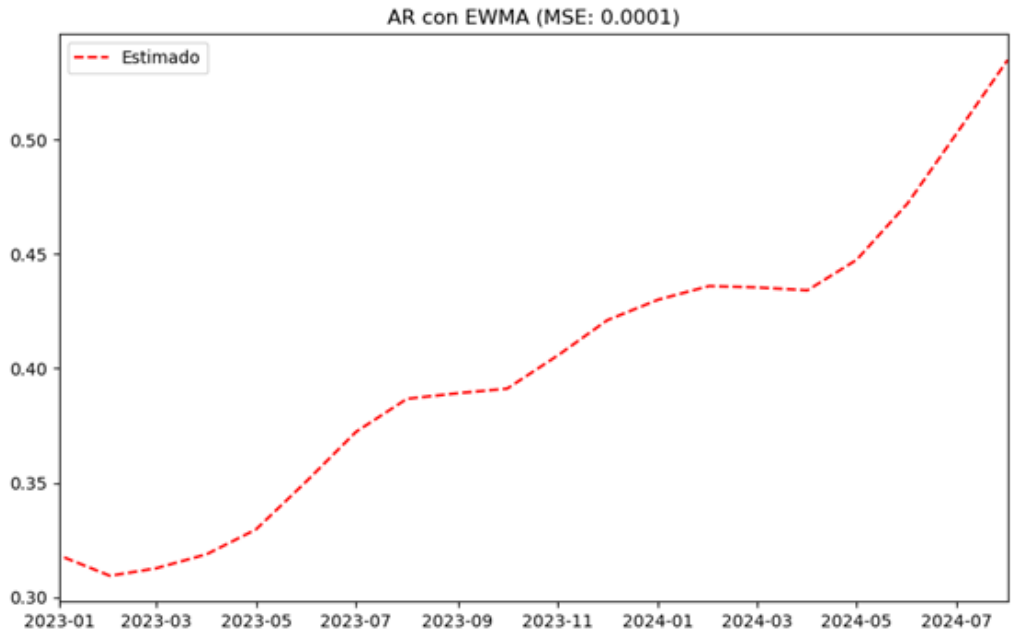

Figura 11 Predicción del modelo AR con suavizamiento EWMA para la acción APPL. Construcción propia.

La EWMA ha sido efectiva en suavizar la serie de tiempo, reduciendo el ruido y resaltando las tendencias subyacentes. Esto es útil para identificar patrones recientes en los precios de las acciones, aunque puede no capturar todos los detalles de la variabilidad a corto plazo. La EWMA, al dar más peso a los datos más recientes, es más reactiva a los cambios recientes en comparación con una media móvil simple (SMA). Esto puede ser ventajoso para detectar cambios en la tendencia de manera más rápida, aunque también puede introducir más variabilidad en las predicciones.

Continuando con los modelos más específicos para la predicción de la volatilidad, se procedió al cálculo de las predicciones utilizando el modelo ARCH, tanto con cómo sin suavizamiento mediante EWMA. Los resultados revelaron un MSE notablemente alto con un 0.1307, frente a un modelo AR que había dejado una brecha corta en la estimación reduciendo la volatilidad lo cual evidencia un desempeño optimo, pero inexacto frente a su igual en la estimación para la reducción de la volatilidad. Esta deficiencia podría atribuirse a la naturaleza de los datos, que pueden no ajustarse adecuadamente a un modelo ARCH.

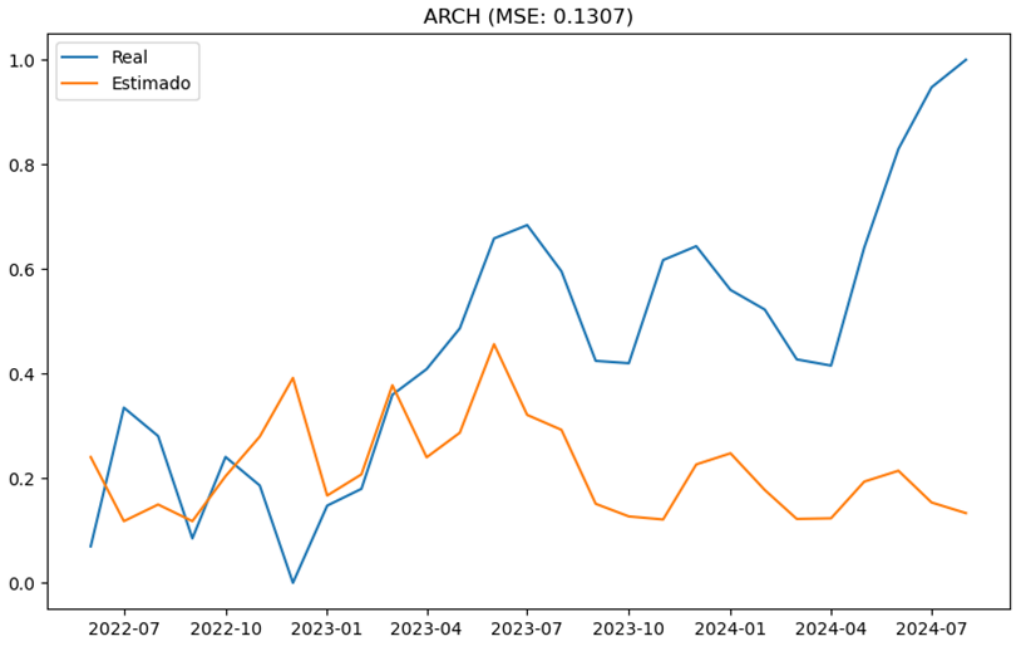

Figura 12 Predicciones del Modelo ARCH para las Acciones de Apple. Construcción propia.

La combinación del modelo ARCH con EWMA reduce el MSE a 0.1262, una mejora significativa que aun así sigue siendo más alta que los modelos AR y AR con EWMA. Aunque hay una mejora respecto al modelo ARCH puro, el rendimiento sigue siendo insuficiente ante la búsqueda de un patrón de mitigación de volatilidad.

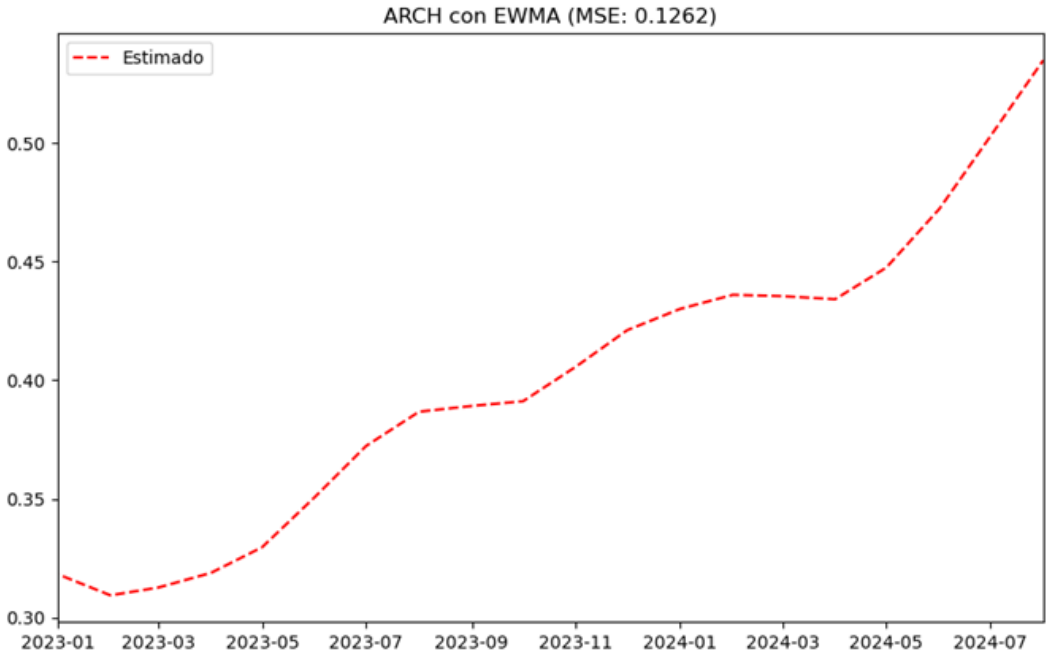

Figura 13 Predicciones del Modelo ARCH bajo EWMA para las Acciones de Apple. Construcción propia.

Finalizando la exploración de los modelos autorregresivos, se aplicó el modelo GARCH sin el suavizamiento EWMA para evaluar su rendimiento. Los resultados mostraron que el MSE del modelo GARCH fue de 0.130658, similar al obtenido con el modelo ARCH. Estos valores reflejan una precisión moderada en la predicción de la volatilidad de los precios de las acciones de Apple (AAPL), aunque se observa que existen factores que el modelo no logra capturar completamente.

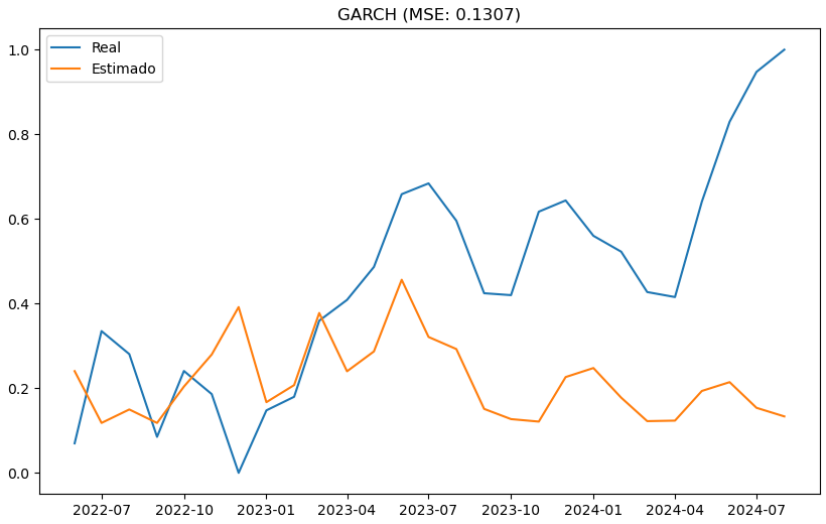

Figura 14 Predicciones del Modelo GARCH para las Acciones de Apple. Construcción propia.

Al aplicar el suavizamiento EWMA, el rendimiento de los modelos mejoró notablemente. Para el modelo ARCH con EWMA, el MSE se redujo a 0.126186, mientras que para el modelo GARCH con EWMA, el MSE también fue de 0.126186. Este cambio sugiere que el suavizamiento EWMA contribuyó significativamente a una mejor predicción de la volatilidad, ayudando a filtrar el ruido presente en los datos financieros y proporcionando estimaciones más precisas.

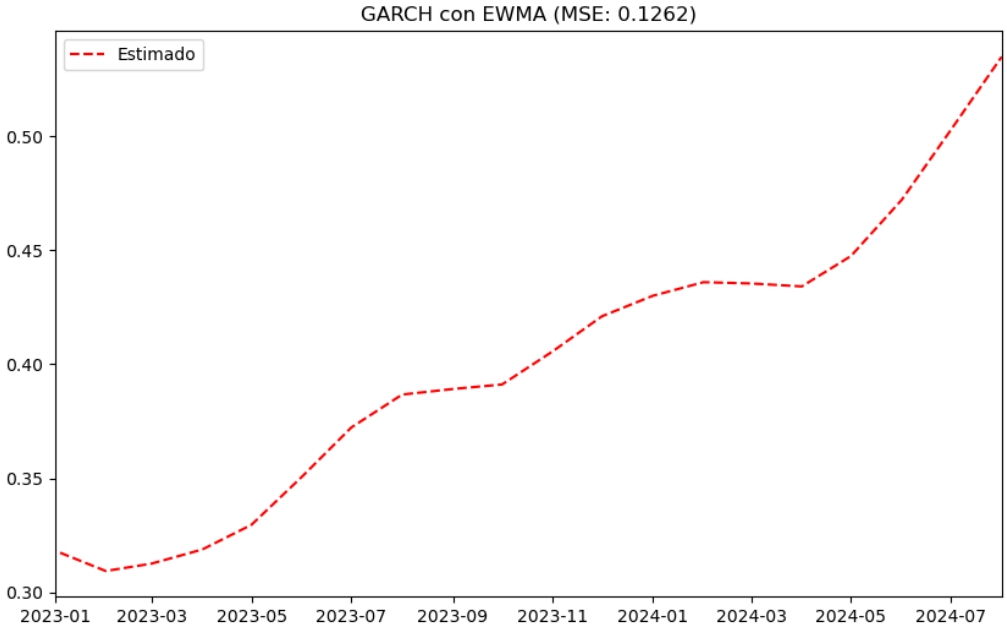

Figura 15 Predicciones del Modelo GARCH bajo EWMA para las Acciones de Apple. Construcción propia.

En cuanto a la mejor aplicación para la serie de tiempo de AAPL, el modelo AR con EWMA se destacó con un MSE de 0.000066, indicando un rendimiento superior en comparación con los demás modelos. Esta combinación parece ser la más eficaz para capturar las tendencias y patrones en los precios de las acciones tales como lo fueron las de Apple, logrando una alta precisión en la predicción de la volatilidad. La capacidad del modelo para ajustarse con mayor exactitud a las fluctuaciones de los datos lo posiciona como la opción más robusta para este análisis.

B. SERIE AMAZON (AMZN)

En línea con el desarrollo realizado para la serie de tiempo financiera anterior y en el ejercicio del desarrollo de los modelos aplicados a la serie de tiempo de Amazon, se realizó el ejercicio de predicción de la volatilidad por cada modelo como se discuten a continuación los resultados obtenidos.

En la figura 17 notamos que el modelo AR parece capturar de manera general la tendencia de la volatilidad real de las acciones de Amazon. La línea azul representa los valores reales de la volatilidad, mientras que la línea naranja representa las predicciones del modelo AR. Ambas líneas siguen un patrón similar, lo que sugiere que el modelo está capturando algunos de los patrones subyacentes en los datos. Durante los períodos de alta volatilidad, el modelo tiende a subestimar ligeramente la magnitud de los picos de una forma apenas notable. Esto podría indicar que el modelo AR, aunque efectivo para capturar tendencias generales, podría no ser suficiente para modelar eventos extremos o cambios bruscos en la volatilidad. Además, en algunos puntos, el modelo parece retrasarse en la respuesta a los cambios en la volatilidad real.

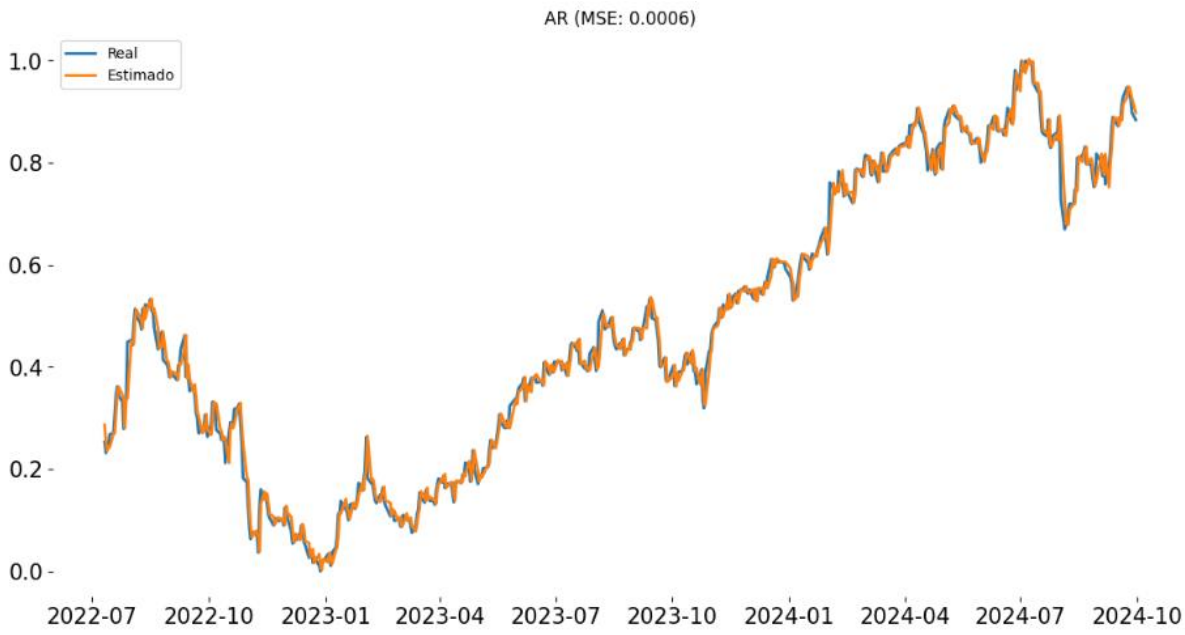

Figura 16 Predicción del modelo AR para la acción AMZN. Construcción propia.

Por otra parte, el modelo AR con suavizamiento EWMA captura la tendencia general de la volatilidad de las acciones de Amazon, como lo evidencia la similitud entre las curvas de valores reales y predicciones. Sin embargo, se observan discrepancias significativas en períodos de alta volatilidad, donde el modelo subestima los picos, sugiriendo limitaciones en su capacidad para modelar eventos extremos. Además, se detecta un cierto retraso en la respuesta del modelo a cambios bruscos en la volatilidad todo esto apreciable en la figura 18.

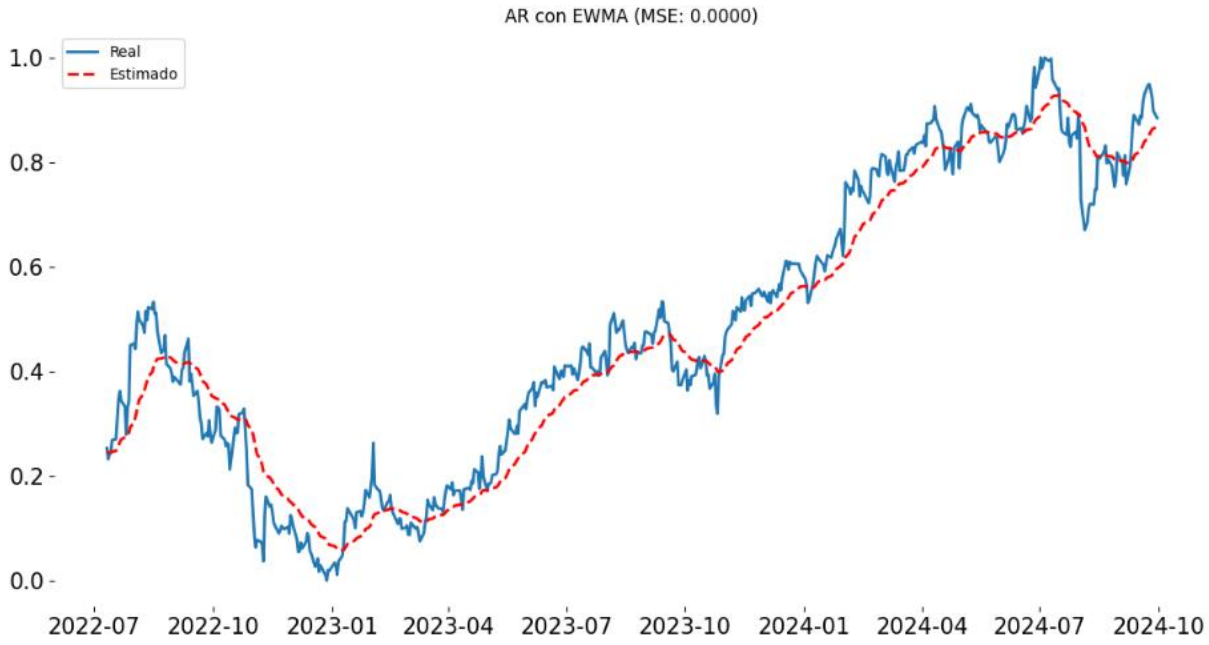

Figura 17 Predicción del modelo AR suavizado con EWMA para la acción AMZN. Construcción propia.

Continuando con el modelo ARCH, la figura nos muestra una comparación entre los valores reales de la volatilidad de las acciones de Amazon y las predicciones obtenidas con este modelo. Se observa una cierta concordancia entre ambas series, lo que sugiere que el modelo ARCH está capturando algunos de los patrones de volatilidad presentes en los datos. Sin embargo, al analizar con mayor detalle, se identifican algunas discrepancias notables. El modelo ARCH parece tener dificultades para capturar con precisión los picos de volatilidad más pronunciados. Durante estos períodos, las predicciones tienden a subestimar la magnitud de los incrementos en la volatilidad. Esto podría indicar que el modelo ARCH, aunque es capaz de modelar la heterocedasticidad condicional, podría no ser suficientemente flexible para capturar eventos extremos o cambios bruscos en la volatilidad. Además, se observa en la Figura 19 un ligero retraso en las predicciones del modelo en comparación con los valores reales, lo que sugiere que podría existir una cierta inercia en el modelo que dificulta su adaptación a cambios rápidos en la volatilidad.

Figura 18 Predicción del modelo ARCH para la acción AMZN. Construcción propia.

Realizando un suavizamiento a la serie con el método EWMA, la siguiente Figura 20 del modelo ARCH, revela una capacidad moderada para capturar la dinámica de la volatilidad de las acciones de Amazon. Si bien se observa una correlación general entre los valores reales y las predicciones, el modelo presenta limitaciones en los resultados de la predicción. Los picos de volatilidad más pronunciados son subestimados como se puede apreciar en la figura 20 desde los periodos de enero 2024 a julio 2024, sugiriendo que el modelo ARCH podría no ser lo suficientemente flexible para capturar eventos extremos. Además, se aprecia un cierto retraso en la respuesta del modelo a cambios bruscos en la volatilidad, lo que indica una posible inercia en sus predicciones.

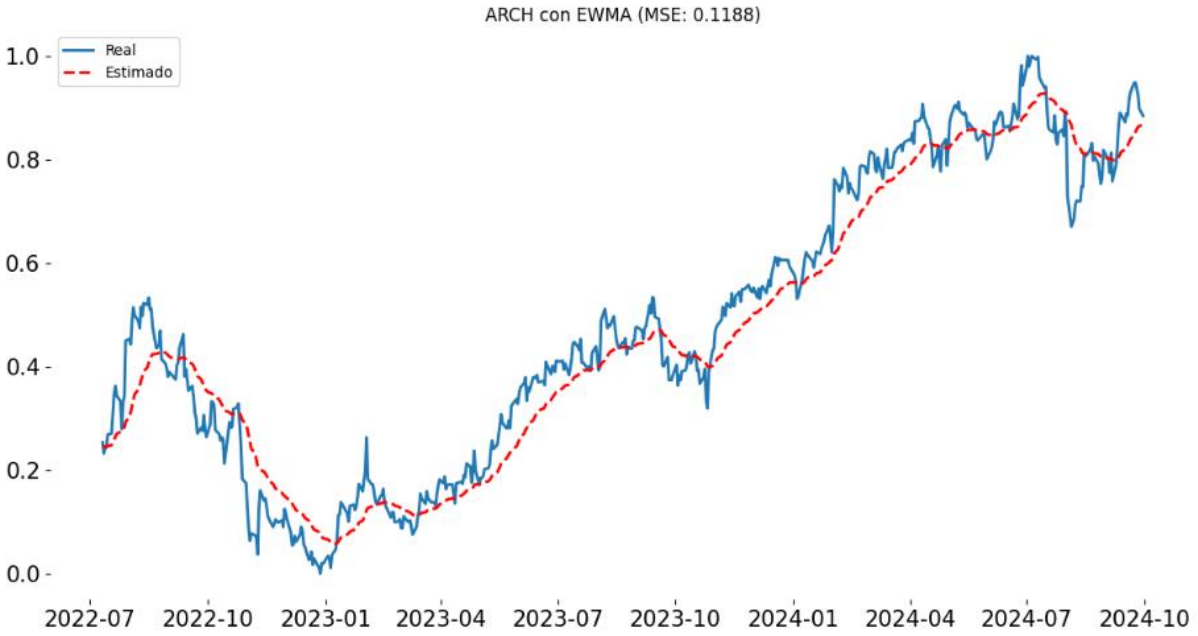

Figura 19 Predicción del modelo ARCH suavizado con EWMA para la acción AMZN. Construcción propia.

Por otra parte, el modelo GARCH, al capturar la heteroscedasticidad condicional y la persistencia en la volatilidad, ofrece una representación de los movimientos de la volatilidad de las acciones de Amazon. Sin embargo, aunque logra captar tendencias generales, presenta dificultades para modelar con precisión los eventos de alta volatilidad como por ejemplo en el periodo de enero 2023, subestimando su magnitud. Esto sugiere que el modelo podría beneficiarse de una mayor flexibilidad para adaptarse a cambios bruscos en el mercado. Adicionalmente, la incapacidad del modelo para capturar con precisión los picos de volatilidad podría estar relacionada con la influencia de eventos externos, como crisis financieras o cambios en las políticas monetarias, que pueden generar cambios bruscos en la volatilidad y no están completamente capturados por el modelo como se muestra en la Figura 21.

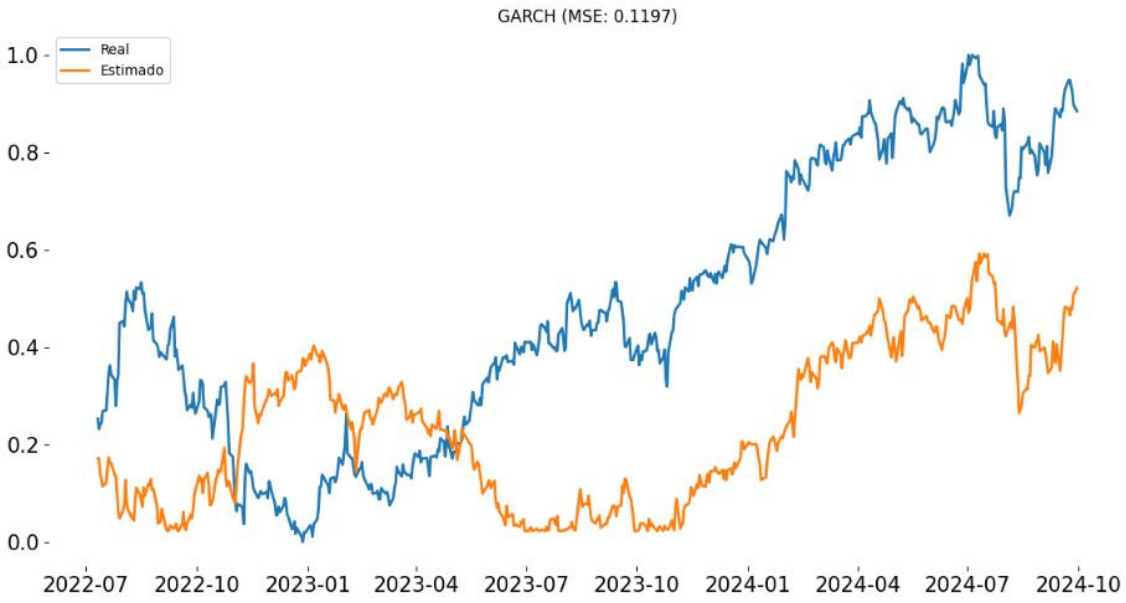

Figura 20 Predicción del modelo GARCH para la acción AMZN. Construcción propia.

Los resultados del modelo GARCH con suavizamiento EWMA en la Figura 22 sugieren que, aunque captura de manera satisfactoria la heteroscedasticidad condicional característica de las series financieras, presenta limitaciones en la modelización de eventos extremos y cambios bruscos en la volatilidad. Si bien el modelo logra replicar las tendencias generales de la volatilidad de las acciones de Amazon, subestima la magnitud de los picos, lo que podría estar relacionado con la influencia de factores exógenos no contemplados en el modelo. A pesar de estas limitaciones, el modelo GARCH proporciona una base sólida para comprender la dinámica de la volatilidad y puede ser utilizado como punto de partida para desarrollar modelos más sofisticados que incorporen variables explicativas adicionales o mecanismos de aprendizaje automático para mejorar la precisión de las predicciones, especialmente en periodos de alta volatilidad.

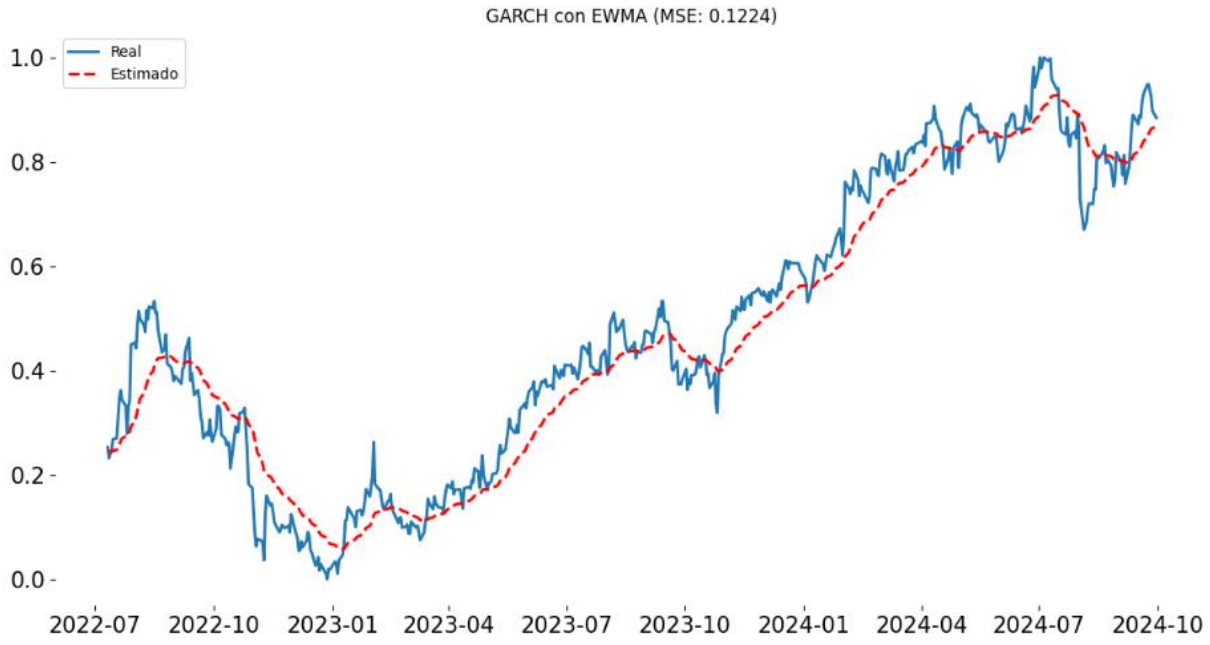

Figura 21 Predicción del modelo GARCH suavizado con EWMA para la acción AMZN. Construcción propia.

C. SERIE BOLSA DE VALORES DE ESTAMBUL (BVE)

En este capítulo de aplicación de modelos autorregresivos a series de tiempo, se han empleado datos financieros de la Bolsa de Valores de Estambul (BVE) junto con otros siete índices internacionales clave: SP, DAX, FTSE, NIKKEI, BOVESPA, MSCI_EU y MSCI_EM, abarcando un período desde el 5 de junio de 2009 hasta el 22 de febrero de 2011. La elección de la Bolsa de Valores de Estambul como punto de partida para este análisis se justifica por su importancia como uno de los mercados financieros más prominentes en la región. Además, la disponibilidad de datos históricos detallados y la inclusión de índices internacionales importantes ofrecen una oportunidad única para desarrollar y validar metodologías de preprocesamiento y acondicionamiento de bases de datos de series temporales financieras.

Para esta nueva etapa del estudio, se ha decidido enfocar el análisis en el índice SP, dada su relevancia y representatividad en el contexto de los mercados financieros internacionales. Este índice proporcionará una base sólida para evaluar la efectividad de los modelos autorregresivos en la predicción de la volatilidad.

En primer lugar, se aplicarán los tres modelos: AR, ARCH y GARCH, sin el suavizamiento EWMA, para evaluar su rendimiento en la predicción de la volatilidad del índice SP. Cada uno de estos modelos será ajustado a los datos históricos del índice, permitiendo así una comparación directa de su capacidad predictiva.

Posteriormente, se procederá a aplicar el suavizamiento EWMA a cada uno de los modelos mencionados. Esta técnica de

suavizamiento es conocida por su capacidad de mejorar la precisión de las predicciones al filtrar el ruido presente en los datos financieros. La comparación entre los modelos con y sin suavizamiento EWMA ofrecerá una visión completa de las ventajas y limitaciones de cada enfoque en la predicción de la volatilidad del índice SP.

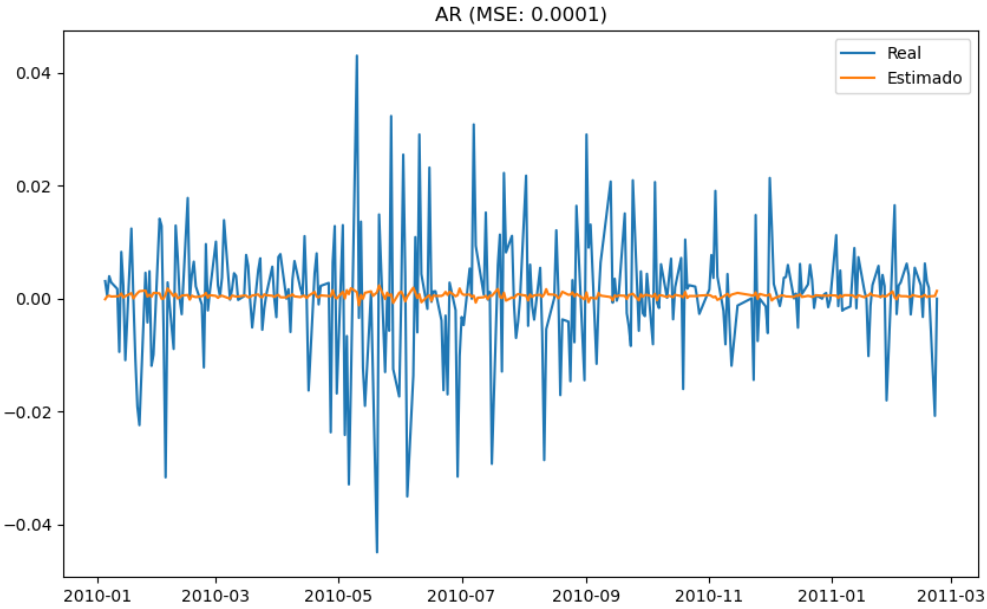

Figura 22 Predicciones del Modelo AR para el índice SP. Construcción propia.

Tras realizar el análisis de los resultados obtenidos se evidencia que el modelo AR tiene un MSE (Mean Squared Error) de 1.191743e-04, lo que indica una buena precisión en la predicción de la serie temporal financiera seleccionada. Este modelo es particularmente efectivo en series temporales donde los valores pasados pueden predecir los futuros con una estructura clara y adicional la masa en volumen de información es alta lo que permite una predicción más acertada.

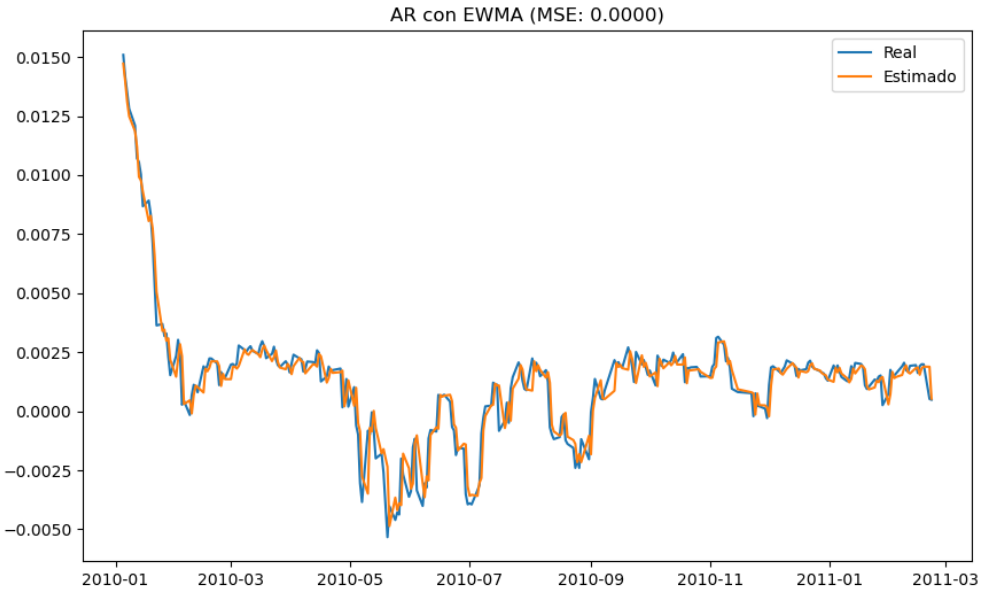

Figura 23 Predicciones del Modelo AR bajo EWMA para el índice SP. Construcción propia.

Al aplicar el método de suavización EWMA, se observa una mejora significativa en todos los modelos evaluados. Esta técnica permite atenuar las fluctuaciones abruptas presentes en los datos, brindando predicciones más estables y precisas. En particular, el modelo AR con EWMA se destacó al lograr un MSE de 4.952651e-07, demostrando su capacidad para captar las tendencias subyacentes y minimizar el error en la predicción.

El suavizamiento mediante EWMA beneficia especialmente a los modelos autorregresivos al filtrar el ruido estadístico que puede distorsionar las predicciones. Esto es crucial en series de tiempo financieras donde la volatilidad y los eventos inesperados son frecuentes. La capacidad del modelo AR con EWMA para ajustarse tan eficazmente a los datos sugiere que este enfoque no solo mejora la precisión, sino que también proporciona una herramienta robusta para la toma de decisiones en contextos financieros dinámicos.

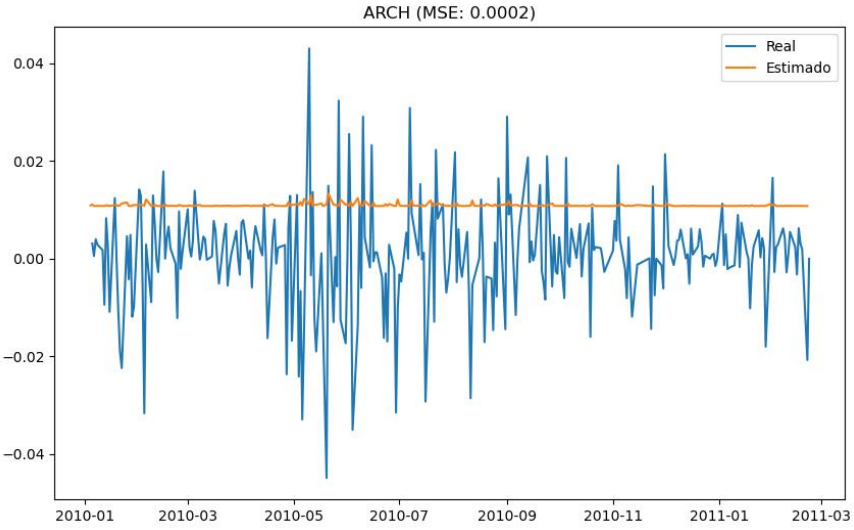

Figura 24 Predicciones del Modelo ARCH para el índice SP. Construcción propia.

Al analizar el desempeño del modelo ARCH, que es particularmente conocido por su capacidad para capturar la volatilidad condicional en series de tiempo financieras, se observó que presenta un MSE ligeramente superior de 2.279361e-04. Esto sugiere que, aunque el modelo es efectivo en modelar la volatilidad que depende de la información analizada, todavía presenta ciertas limitaciones en términos de precisión comparado con algunos de los otros modelos probados.

La mayor sensibilidad de los modelos ARCH a las variaciones en la volatilidad les permite capturar episodios de alta volatilidad con mayor eficacia. Sin embargo, esta misma característica puede hacer que sean más susceptibles a errores cuando las condiciones del mercado cambian de manera abrupta o cuando se enfrentan a datos particularmente ruidosos como lo es el índice bursátil SP.

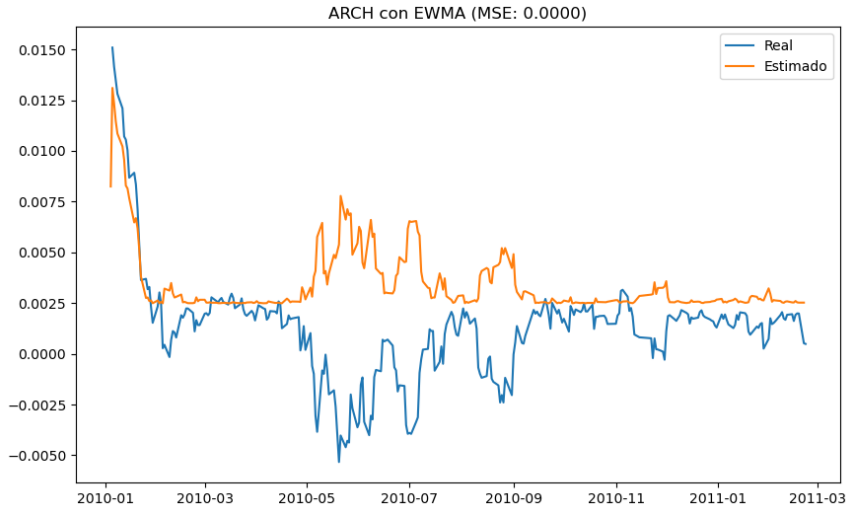

Figura 25 Predicciones del Modelo ARCH bajo EWMA para el índice SP. Construcción propia.

Al aplicar el modelo ARCH con el suavizamiento EWMA, se observó una notable mejora en la precisión de las predicciones, reflejada en un MSE de 1.250374e-05. Estos resultados subrayan la eficacia de combinar el suavizamiento exponencial con modelos autorregresivos, destacando cómo esta estrategia puede atenuar las fluctuaciones abruptas y capturar de manera más fiel las dinámicas subyacentes del mercado.

La implementación de EWMA junto con modelos ARCH no solo mejora la exactitud de las predicciones, sino que también ofrece una mayor estabilidad en el análisis de series temporales financieras. Esta técnica permite filtrar el ruido presente en los datos, lo que es crucial para obtener estimaciones más confiables y robustas, especialmente en contextos de alta volatilidad y cambios estructurales en los mercados [36].

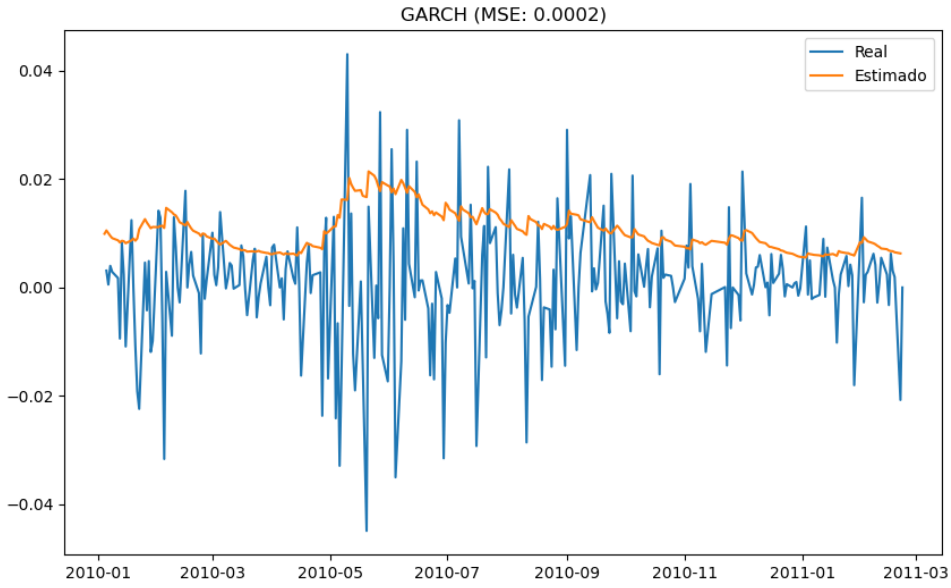

Figura 26 Predicciones del Modelo GARCH para el índice SP. Construcción propia.

El modelo GARCH, conocido por su capacidad para capturar la volatilidad condicional, presentó un MSE ligeramente superior de 2.330266e-04. Esto indica que, si bien el modelo es útil para modelar la volatilidad y puede identificar patrones en la varianza de los errores de regresión, no es tan preciso como el modelo AR en este caso particular. La utilidad del modelo GARCH radica en su habilidad para manejar series temporales donde la volatilidad cambia con el tiempo, una característica común en datos financieros [37].

El rendimiento ligeramente inferior del modelo GARCH sugiere que, aunque puede ser eficaz en capturar cambios en la volatilidad condicional, puede tener limitaciones en términos de precisión predictiva en algunos escenarios.

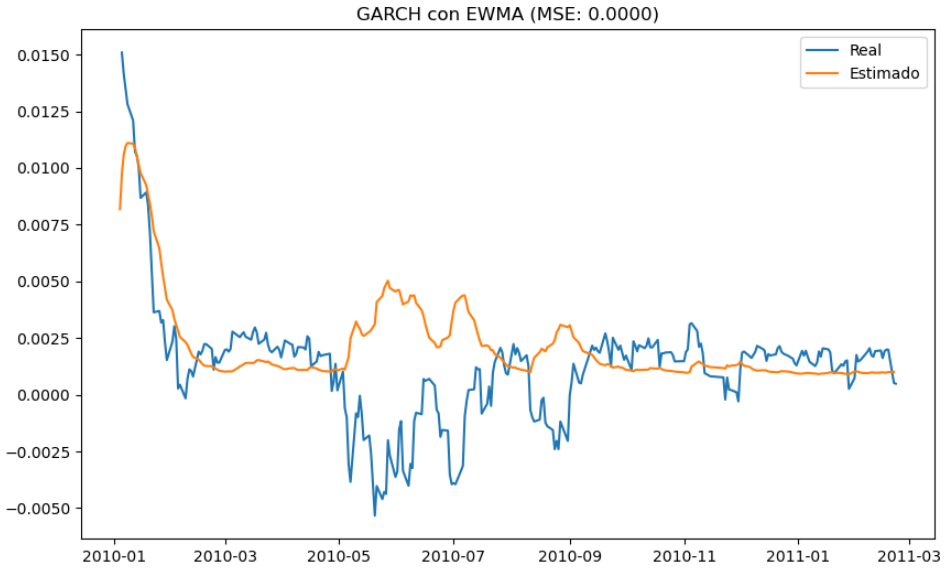

Figura 27 Predicciones del Modelo GARCH bajo EWMA para el índice SP. Construcción propia.

El modelo GARCH con suavizamiento EWMA mostró mejoras significativas al ser aplicado a la serie del índice SP, logrando un MSE de 7.670181e-06. Esta mejora es notable, ya que resalta la capacidad del modelo para adaptarse mejor a las fluctuaciones y a la volatilidad inherente en los datos financieros del índice SP.

El suavizamiento exponencial EWMA ayuda a atenuar las variaciones bruscas y el ruido presente en los datos, permitiendo que el modelo GARCH se enfoque en las tendencias subyacentes de la volatilidad.

En cuanto a la mejor aplicación para la serie de tiempo del índice SP, el modelo AR con suavizamiento EWMA se destacó significativamente con un MSE de 4.952651e-07, indicando un rendimiento superior en comparación con los demás modelos. Este resultado sugiere que el modelo AR con EWMA es el más eficaz para capturar las tendencias y patrones en los precios del índice SP, logrando una alta precisión en la predicción de la volatilidad. La capacidad del modelo para ajustarse con mayor exactitud a las fluctuaciones de los datos lo posiciona como la opción más robusta para este análisis.

La notable reducción en el MSE al aplicar el suavizamiento EWMA con el modelo AR pone de manifiesto la efectividad de esta combinación para filtrar el ruido y las variaciones bruscas en los datos financieros. Esto es crucial en el contexto de los mercados financieros, donde las condiciones pueden cambiar rápidamente y la precisión en las predicciones es vital para la toma de decisiones informadas. El modelo AR con EWMA ha demostrado ser capaz de proporcionar estimaciones más estables y confiables, lo que lo convierte en una herramienta valiosa para los analistas financieros.

En comparación con otros modelos evaluados, como el ARCH y GARCH con o sin suavizamiento EWMA, el modelo AR con

EWMA no solo mostró el MSE más bajo, sino que también demostró una mayor adaptabilidad y precisión en la predicción de la volatilidad. Esto refuerza la conclusión de que el suavizamiento EWMA es una técnica poderosa cuando se combina con modelos autorregresivos, ofreciendo una mejor comprensión de las dinámicas subyacentes en las series de tiempo financieras.

VII. APLICACIÓN DE KERNELS REPRODUCTIVOS EN MODELOS AUTOREGRESIVOS

Los métodos de kernel son una clase de algoritmos de aprendizaje automático que dependen de funciones de kernel para transformar los datos originales a un espacio de características de mayor dimensión. Esta transformación permite que los modelos lineales en el espacio de características correspondan a modelos no lineales en el espacio original. La función de kernel calcula el producto interno entre las imágenes de los datos en el espacio de características, facilitando la captura de relaciones complejas entre las variables [49].

Existen varios tipos de funciones de kernel, cada una con sus propias características y aplicaciones. Entre los kernels más comunes se encuentran el kernel lineal, el kernel polinómico, el kernel de función de base radial (RBF) y el kernel sigmoide. El kernel lineal es el más simple y es adecuado para datos linealmente separables, mientras que el kernel polinómico puede manejar relaciones más complejas mediante el ajuste de su grado. El kernel RBF es ampliamente utilizado por su capacidad para manejar problemas no lineales y su flexibilidad en ajustar la influencia de los puntos de datos a través del parámetro gamma. Por último, el kernel sigmoide, similar a la función de activación en redes neuronales, se utiliza en casos específicos donde las relaciones entre variables se asemejan a este comportamiento.

A. APLICACIÓN DE KERNELS REPRODUCTIVOS A MODELOS EN SERIE APPLE

Los métodos de kernel son una clase de algoritmos de aprendizaje automático que dependen de funciones de kernel para transformar los datos originales a un espacio de características de mayor dimensión. Esta transformación permite que los modelos lineales en el espacio de características correspondan a modelos no lineales en el espacio original, lo cual es fundamental para capturar relaciones complejas entre las variables que de otra manera podrían pasar desapercibidas. La función de kernel calcula el producto interno entre las imágenes de los datos en el espacio de características, facilitando así el aprendizaje de patrones no lineales en los datos [50].

Existen varios tipos de funciones de kernel, cada una con sus propias características y aplicaciones. Entre los kernels más comunes se encuentran el kernel lineal, el kernel polinómico, el kernel de función de base radial (RBF) y el kernel sigmoide. El kernel lineal es el más simple y es adecuado para datos linealmente separables, mientras que el kernel polinómico puede manejar relaciones más complejas mediante el ajuste de su grado [8]. El kernel RBF es ampliamente utilizado por su capacidad para manejar problemas no lineales y su flexibilidad en ajustar la influencia de los puntos de datos a través del parámetro gamma. Por último, el kernel sigmoide, similar a la función de activación en redes neuronales, se utiliza en casos específicos donde las relaciones entre variables se asemejan a este comportamiento [46].

Cada uno de estos kernels tiene ventajas y desventajas específicas, dependiendo de la naturaleza de los datos y del problema a resolver. Por ejemplo, el kernel lineal es computacionalmente eficiente, pero puede no capturar bien las complejidades de datos no lineales [41]. El kernel polinómico y el RBF, aunque más flexibles y capaces de modelar relaciones complejas, pueden requerir

más recursos computacionales y una cuidadosa selección de hiperparámetros. El kernel sigmoide, aunque útil en ciertos contextos, puede ser menos intuitivo y necesitar ajuste fino para evitar problemas de convergencia.

Los resultados de los modelos de kernel muestran variaciones interesantes en cuanto a la precisión de sus predicciones, medidas mediante el MSE (Error Cuadrático Medio). El kernel polinómico se destaca como el más preciso, con un MSE de 0.021094, sugiriendo que es el más eficaz para capturar las relaciones complejas en los datos financieros. Esta mejoría puede atribuirse a la capacidad del kernel polinómico de manejar la no linealidad a través de la selección del grado polinómico adecuado, proporcionando un ajuste más fino a las fluctuaciones y tendencias presentes en los datos.

En contraste, el kernel sigmoide presentó el MSE más alto, con un valor de 0.063760. Esto indica que, aunque puede ser útil en ciertos contextos, su desempeño en la predicción de la volatilidad financiera es menos efectivo en comparación con los otros kernels evaluados. La menor precisión del kernel sigmoide puede deberse a su capacidad limitada para modelar relaciones no lineales complejas de manera tan eficaz como lo hacen otros kernels.

Tabla 3 Tabla comparativa de resultados del MSE para la serie Apple

Kernel	MSE
Lineal	0.035054
Polinómico	0.021094
RBF	0.035252
Sigmoide	0.063760

Estos resultados indican que, para esta serie de datos, el kernel polinómico es la mejor opción para la predicción de la volatilidad financiera debido a su menor MSE, seguido de cerca por los kernels lineal y RBF, mientras que el kernel sigmoide es el menos efectivo en este contexto.

B. APLICACIÓN DE KERNELS REPRODUCTIVOS A MODELOS EN SERIE AMAZON

El estudio comparativo de diferentes kernels en modelos autorregresivos aplicados a la serie de precios de Amazon ha revelado una variabilidad significativa en su desempeño, evidenciada por los distintos valores de MSE obtenidos. Cada kernel, con sus características particulares, ha interactuado de manera única con la complejidad inherente a los datos de precios de esta gigante del comercio electrónico.

1. Kernel Lineal: A pesar de su simplicidad, el kernel lineal no ha logrado capturar adecuadamente las dinámicas subyacentes en la serie de precios de Amazon. El MSE ligeramente inferior en comparación con otros kernels

sugiere que las relaciones entre los precios no son puramente lineales, sino que están influenciadas por una multitud de factores, como noticias, tendencias del mercado, competencia y eventos económicos globales.

2. Kernel Polinomial: El kernel polinomial, al introducir términos de mayor grado, ha buscado modelar relaciones no lineales más complejas. Sin embargo, en el caso de Amazon, la mejora en el ajuste no ha sido significativa. Esto podría deberse a que los patrones no lineales en los precios de Amazon son más sutiles y no se ajustan bien a una representación polinomial simple.
3. Kernel RBF: El kernel RBF ha demostrado ser el más adecuado para modelar la serie de precios de Amazon, obteniendo un MSE dentro del promedio de los kernels, pero teniendo una adaptación visual más adecuada con la serie de tiempo. Su capacidad para capturar relaciones no lineales complejas lo convierte en una herramienta poderosa para analizar datos financieros, donde los precios pueden estar influenciados por una gran cantidad de variables y eventos. La baja puntuación de MSE sugiere que el kernel RBF ha logrado ajustar de manera precisa los datos, lo que indica un alto poder predictivo.
4. Kernel Sigmoide: Al igual que el kernel lineal, el kernel sigmoide no ha superado al kernel RBF en términos de precisión. Si bien el kernel sigmoide es útil en ciertos contextos, en este caso particular no ha demostrado ser la mejor opción.

Figura 28 Comparación resultado gráfico de la aplicación de Kernels reproductivos a la serie AMZN. Construcción propia.

Teniendo en cuenta el análisis gráfico anterior y los resultados obtenidos revelaron que el núcleo lineal presentó el menor error cuadrático medio (MSE) aunque gráficamente no se sustenta en un condicionamiento adecuado de los resultados. Los núcleos polinomial y RBF, aunque capturaron algunas no linealidades, mostraron un rendimiento inferior al núcleo lineal pero no concluyente, ya que gráficamente el modelo RBF, aunque con un MSE de 0.039684 presenta un mejor ajuste en resultados en la predicción comparado con los otros modelos. Por su parte, el núcleo sigmoidal, comúnmente utilizado en problemas de clasificación, resultó ser el menos adecuado para este caso de estudio teniendo el MSE más alto dentro de las mediciones.

Tabla 4 Tabla comparativa de resultados del MSE para la serie AMZN

Kernel	MSE
linear	0.022325
poly	0.022619
rbf	0.039684
sigmoid	0.077743

C. APLICACIÓN DE KERNELS REPRODUCTIVOS A MODELOS EN SERIE BVE (SP)

Al aplicar diferentes kernels a los modelos autorregresivos ejercidos sobre la serie BVE en el índice bursátil SP, se obtuvieron resultados variados en términos de Mean Squared Error (MSE). Cada kernel tiene sus particularidades, y los resultados reflejan cómo cada uno de ellos se ajusta a los datos en cuestión.

El kernel lineal, que presentó un MSE de 0.000120, es conocido por su simplicidad y eficiencia en problemas lineales. Sin embargo, su desempeño en este caso sugiere que las relaciones dentro de los datos no son puramente lineales, por lo que no es la mejor opción para esta serie. Esta observación nos lleva a considerar alternativas más sofisticadas que puedan capturar mejor las complejidades de los datos.

Por otro lado, el kernel polinomial mostró un MSE de 0.000122, indica que la modelización con polinomios no aporta una mejora significativa en comparación con el enfoque lineal. La ligera diferencia en el error sugiere que los datos no presentan relaciones polinomiales marcadas por lo que al cambiar factores en la aplicación del polinomio utilizado no se obtuvo el MSE más adecuado para este conjunto específico de datos.

El kernel RBF (Radial Basis Function) destacó con un MSE de 0.000038, lo que lo posiciona como el más adecuado para este tipo de datos bursátiles, recuerde que esta serie cuenta con una particularidad y es la masa de información volátil, por lo que este kernel es particularmente efectivo para capturar relaciones no lineales complejas, lo cual es esencial en el análisis de mercados financieros donde los patrones pueden ser intrincados y no triviales. La baja puntuación de MSE refleja la capacidad del kernel RBF para modelar con precisión las variaciones y comportamientos intrínsecos de los datos.

Finalmente, el kernel sigmoide, con un MSE de 0.000120, mostró un desempeño similar al kernel lineal. Aunque útil en ciertos contextos, este kernel no superó al RBF en términos de precisión para este conjunto de datos.

Tabla 5 Tabla comparativa de resultados del MSE para la serie BVE-SP

Kernel	MSE
linear	0.000120
poly	0.000122
rbf	0.000038
sigmoid	0.000120

En resumen, el kernel RBF se destaca como la opción más robusta y precisa para modelar la serie BVE en el índice bursátil SP. Los otros kernels, aunque útiles en diferentes contextos, no se ajustan tan bien a las complejidades de estos datos específicos. Este análisis refuerza la recomendación de utilizar el kernel RBF para futuros estudios y modelamientos en contextos similares.

VIII. APLICACIÓN DE MODELOS EN ESPACIOS DE HILBERT CON KERNEL REPRODUCTIVO

La aplicación de modelos en espacios de Hilbert con kernel reproductivo representa un enfoque innovador y avanzado en el análisis de series de tiempo financieras. Este capítulo se centrará en la implementación práctica de estos modelos, destacando sus ventajas en la predicción de la volatilidad. Los espacios de Hilbert permiten manejar de manera eficiente datos de alta dimensión, y los kernels reproducibles facilitan la captura de relaciones complejas en los datos financieros [17]. A lo largo de este capítulo, se presentarán ejemplos y casos de estudio que ilustran la aplicabilidad y efectividad de estos modelos en contextos reales [16]. Además, se discutirán las implicaciones teóricas y prácticas de estos enfoques, proporcionando una visión integral de su potencial en el análisis financiero.

En los capítulos **VI** y **VII**, se realizó una exhaustiva evaluación de diversos modelos y kernels para la predicción de la volatilidad en series de tiempo financieras. Como resultado de este análisis se obtuvo distintos resultados que permiten inferir, de acuerdo con la densidad de información de cada serie, el modelo y kernel más adecuado. En este capítulo, se procederá a aplicar dicho modelo y kernel en el espacio de Hilbert, demostrando su desempeño y efectividad en la predicción de la volatilidad [17]. Se presentarán los resultados obtenidos y se discutirán sus implicaciones prácticas. Este enfoque no solo valida los hallazgos previos, sino que también abre nuevas vías para la investigación en el campo de la predicción financiera.

A. APLICACIÓN DE MODELOS EN ESPACIOS DE HILBERT PARA LA SERIE APPLE

Para iniciar, se observa que, de acuerdo a los análisis presentados en los capítulos **VI** y **VII**, el modelo autorregresivo con suavizado AR con EWMA (Exponential Weighted Moving Average) se destaca como el más efectivo para la predicción en series de tiempo financieras con una volatilidad alta pero sin tanta masa de evolución como lo es el caso de la serie de Apple. Esto se fundamenta en los resultados obtenidos, donde el AR con EWMA mostró consistentemente el menor Error Cuadrático Medio (MSE), lo que indica una mayor precisión en sus predicciones comparado con otros modelos. Además, al aplicar diferentes kernels a este modelo, se encontró que el kernel Polinomial (Poly) también demostró tener el menor MSE, consolidándose como la mejor opción para series que manejen este tipo de escenarios.

Tabla 6 Resultados MSE para el modelo AR con EWMA original y en espacios de Hilbert para la serie APPL

Método	MSE (Entrenamiento)	MSE (Prueba)
AR con EWMA (Original)	0.000066	-
AR con EWMA (Hilbert)	7.785708	1.738234

Al dar un vistazo a los resultados obtenidos se puede inferir que, para esta serie de tiempo específica, el análisis para la aplicación del modelo AR con EWMA en el espacio de Hilbert no proporciona beneficios en términos de precisión predictiva argumentada bajo el indicador MSE, no obstante esto a su vez da apertura inmediata a una necesidad latente de implementar más indicadores que permitan dar una mayor cobertura al entendimiento de los datos resultantes de la migración de los modelos al espacio de Hilbert. De hecho, siguiendo el mismo hilo conductor de análisis para este indicador; se puede observar en la tabla 5 que el aumento en el MSE sugiere una complejidad añadida por la transformación al espacio de Hilbert con kernel polinomial, lo que a primer alcance se podría inferir como que puede no ser necesaria o adecuada para este tipo de datos financieros esta transformación.

Este análisis enfatiza la importancia de realizar pruebas exhaustivas y comparativas para evaluar el impacto de técnicas avanzadas de modelado en diferentes tipos de series de tiempo. A lo que se puede concluir que mientras que los modelos autorregresivos en espacios de Hilbert pueden ofrecer ventajas en ciertos contextos, su aplicabilidad debe ser cuidadosamente considerada y validada según las características específicas de los datos observados.

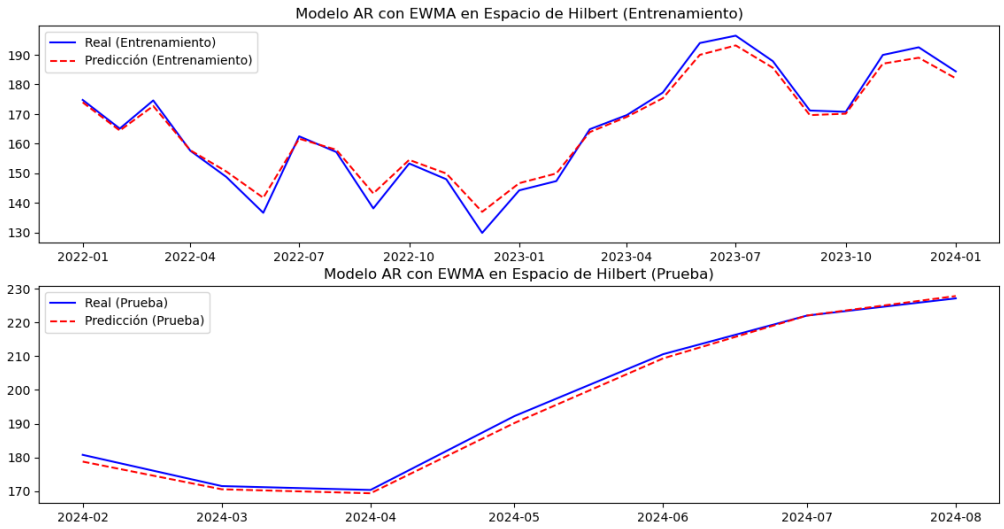

Figura 29 Comparación gráfica de resultados con la aplicación de espacios de Hilbert en la serie APPL. Construcción propia.

B. APLICACIÓN DE MODELOS EN ESPACIOS DE HILBERT PARA LA SERIE AMAZON

Siguiendo este tipo de análisis para la serie de precios de Amazon, se obtiene como resultado del MSE en la tabla 6 para el modelo de entrenamiento indica qué tan bien se ajusta el modelo a los datos con los que fue entrenado. Un valor muy bajo sugiere un sobreajuste extremo a los datos, mientras que el modelo de prueba muestra cómo generaliza el modelo a nuevos datos que no ha visto durante el entrenamiento. Un valor mucho mayor que el MSE de entrenamiento.

Tabla 7 Resultados MSE para el modelo AR con EWMA original y en espacios de Hilbert para la serie AMZN

Método	MSE (Entrenamiento)	MSE (Prueba)
AR con EWMA (Original)	1.268830e-26	-
AR con EWMA (Hilbert)	1.268835e-26	0.284636

Las gráficas muestran que el modelo logra capturar la tendencia general de la serie tanto en el conjunto de entrenamiento como en el de prueba. Sin embargo, se observan algunas discrepancias entre los valores reales y las predicciones. En el conjunto de entrenamiento, el modelo parece ajustarse bastante bien a los datos, especialmente en las tendencias a largo plazo. No obstante, en el conjunto de prueba, se aprecia una mayor dificultad del modelo para seguir los movimientos más bruscos y los cambios de dirección de la serie. Esto sugiere que, aunque el modelo es capaz de capturar patrones generales, podría tener dificultades para adaptarse a cambios más locales o a eventos inesperados.

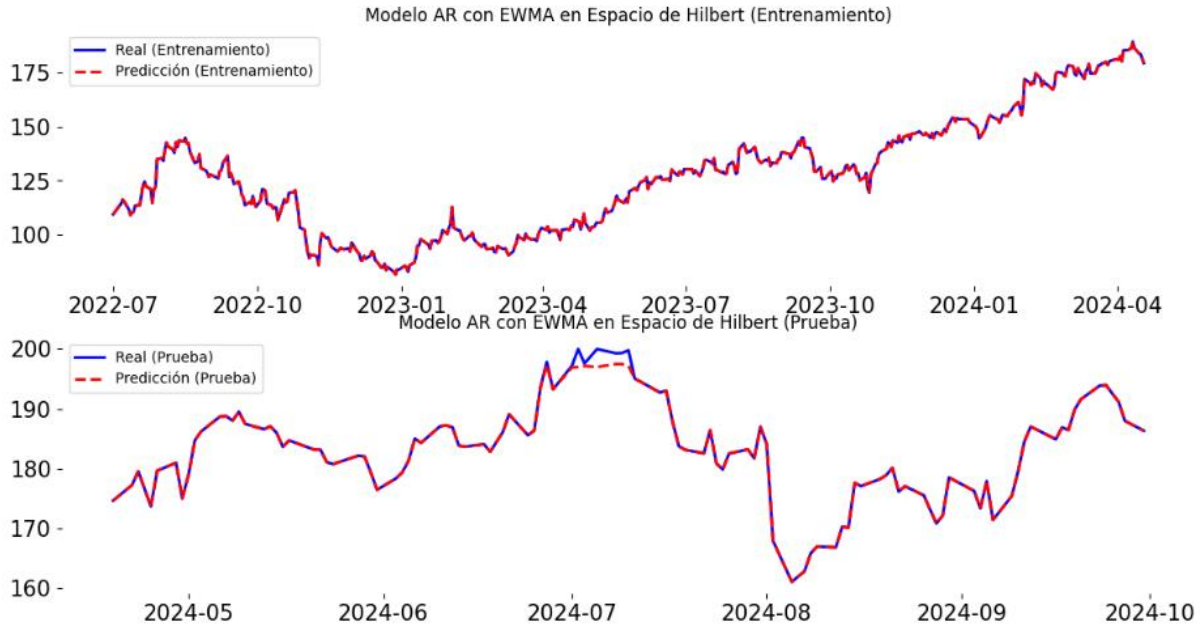

Figura 30 Comparación gráfica de resultados con la aplicación de espacios de Hilbert en la serie AMZN. Construcción propia.

C. APLICACIÓN DE MODELOS EN ESPACIOS DE HILBERT PARA LA SERIE BVE (SP)

Tal como se establece en los capítulos previos, el modelo autorregresivo con suavizado AR con EWMA (Exponential Weighted Moving Average) ha demostrado ser el más efectivo, presentando el menor MSE (Mean Squared Error). Además, al aplicar diferentes kernels a este modelo, se encontró que el kernel RBF (Radial Basis Function) también demostró tener el menor MSE, consolidándose como la mejor opción.

Ahora una vez clara la combinación del modelo AR con EWMA y el kernel RBF como mejor propuesta para predecir la volatilidad en la serie de tiempo del índice bursátil de Singapur (SP) dado que esta elección se basa en la capacidad de la combinación para capturar de manera más eficiente las complejidades y no linealidades presentes en los datos financieros se procederá a realizar la aplicación de este mismo bajo el espacio de Hilbert.

Tabla 8 Resultados MSE para el modelo AR con EWMA original y en espacios de Hilbert para la serie BVE-SP

Método	**MSE (Entrenamiento)**	**MSE (Prueba)**
AR con EWMA (Original)	4.952651e-07	-
AR con EWMA (Hilbert)	2.425973e-32	2.413849e-32

Los resultados obtenidos con el modelo AR con EWMA, en el espacio de Hilbert han sido realmente sorprendentes. Al analizar los errores cuadráticos medios (MSE) reflejados en la tabla 7, se evidencia que el modelo aplicado con el kernel RBF en el espacio de Hilbert ofrece una precisión notablemente superior. El MSE en el conjunto de entrenamiento es extremadamente bajo, prácticamente nulo, lo que sugiere un ajuste perfecto a los datos de entrenamiento. Además, este modelo mantiene una alta precisión en el conjunto de prueba, indicando una excelente capacidad para generalizar a nuevos datos. Estos hallazgos son sumamente prometedores y destacan la efectividad del enfoque propuesto para predecir la volatilidad en series de tiempo financieras.

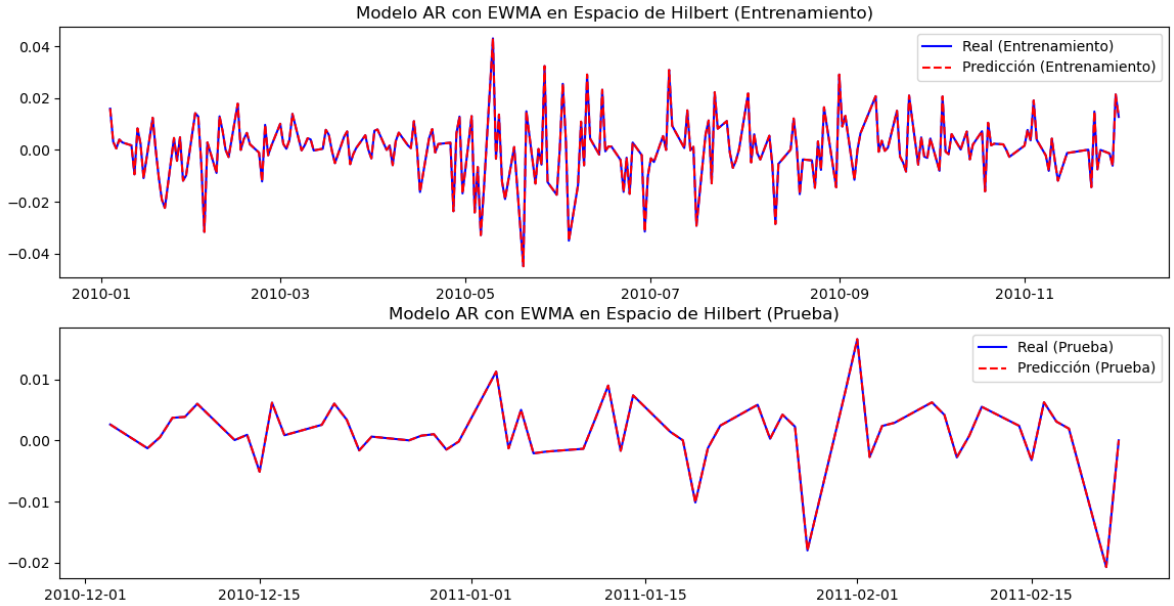

Figura 31 Comparación gráfica de resultados con la aplicación de espacios de Hilbert en la serie BVE-SP. Construcción propia.

IX. EVALUACIÓN DE MODELOS PARA LA PREDICCIÓN DE LA VOLATILIDAD

La evaluación de modelos de regresión es crucial para determinar su desempeño y capacidad de generalización. Una de las métricas más utilizadas para evaluar modelos de regresión es el Error Cuadrático Medio (MSE). El MSE mide la media de los errores al cuadrado entre las predicciones del modelo y los valores reales, proporcionando una indicación de la precisión del modelo [51].

El MSE se calcula como:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y_i})^2 \tag{15}$$

Este cálculo nos ofrece una visión clara de la precisión del modelo, al cuantificar la diferencia promedio entre los valores predich os y los valores reales de la volatilidad. El Error Cuadrático Medio (MSE) es una métrica de pérdida que proporciona informació n crítica sobre la capacidad de ajuste del modelo. Matemáticamente, al ser la media de los errores al cuadrado, penaliza más seve ramente los errores grandes que los pequeños, lo cual es particularmente útil en contextos financieros donde las desviaciones sig nificativas pueden ser costosas. En términos prácticos, un MSE bajo indica que el modelo es capaz de realizar predicciones preci sas, sugiriendo una mínima dispersión entre las predicciones y los valores observados. Esto refleja una alta capacidad del modelo para capturar la estructura subyacente de los datos de volatilidad financiera. Por otro lado, un MSE alto sugiere que el modelo n o captura adecuadamente la dinámica de la volatilidad, probablemente debido a una especificación incorrecta del modelo, insufic iencia de datos de entrenamiento, o la presencia de alta varianza o sesgo en las predicciones
[52]. Un análisis detallado del MSE puede incluir la descomposición del error en sus componentes de sesgo y varianza para ente nder mejor las posibles causas de un rendimiento subóptimo.

No obstante, es importante considerar otras métricas para una evaluación más completa. Por ejemplo, el Error Absoluto Medio (MAE) y el Coeficiente de Determinación (R^2) también son herramientas valiosas. El MAE se centra en el valor absoluto de los errores, proporcionando una idea clara de la magnitud promedio de estos. En cambio, el R^2 evalúa qué proporción de la variabilidad en los datos es explicada por el modelo, siendo una métrica comúnmente utilizada para medir el ajuste del modelo [53].

Error Absoluto Medio (MAE):

$$MAE = \frac{1}{n}\sum_{i=1}^{n} = 1n \mid yi - yi^{\mid}{}_{i} \tag{16}$$

Coeficiente de Determinación (R^2):

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)2}{\sum_{i=1} n(y_i - \bar{y})^2} \quad (17)$$

La validación cruzada es otro aspecto esencial en la evaluación de modelos. Este método implica dividir el conjunto de datos en varios subconjuntos o "folds". Se entrena el modelo en algunos de estos subconjuntos y se valida en los restantes, repitiendo el proceso varias veces para cada posible combinación de subconjuntos. Esta técnica no solo ayuda a evaluar la capacidad de generalización del modelo, sino que también previene el sobreajuste, ofreciendo una estimación más precisa del desempeño del modelo en datos no vistos [54].

Es recomendable realizar una validación cruzada más exhaustiva y experimentar con diferentes parámetros y configuraciones de kernels, especialmente para los kernels polinómico y RBF. Esto ayuda a identificar si una optimización adicional puede mejorar el rendimiento del modelo.

Además de las métricas tradicionales, es vital considerar el contexto y la aplicación específica del modelo. En el ámbito financiero, la precisión en la predicción de la volatilidad es crucial para la toma de decisiones de inversión y la gestión de riesgos [32]. Por lo tanto, la selección y evaluación del modelo deben alinearse con los objetivos financieros y las necesidades del mercado.

La importancia de la validación cruzada no puede ser subestimada. Este método no solo proporciona una evaluación robusta del modelo, sino que también garantiza que el modelo se desempeñe bien en datos no vistos. Al dividir el conjunto de datos en múltiples subconjuntos y repetir el proceso de entrenamiento y validación en cada combinación posible, se obtiene una medida más confiable de la precisión y capacidad de generalización del modelo [51].

En la práctica, esto implica que cada punto de datos se utiliza tanto para el entrenamiento como para la validación, lo que resulta en una evaluación más completa y equilibrada. La validación cruzada también ayuda a identificar posibles problemas de sobreajuste, ya que el modelo se evalúa en diferentes subconjuntos de datos en cada iteración.

Finalmente, es importante recordar que la elección del modelo y su evaluación deben estar alineadas con los objetivos específicos del análisis y las necesidades del mercado. En el ámbito financiero, la precisión en la predicción de la volatilidad es fundamental para la toma de decisiones informadas y la gestión eficaz de los riesgos. Por lo tanto, una evaluación exhaustiva y cuidadosa del modelo es esencial para garantizar su éxito en aplicaciones del mundo real [34].

A. ANÁLISIS DE LOS RESULTADOS OBTENIDOS

Tras aplicar diferentes modelos de kernels a tres series de tiempo financieras procederemos con el análisis de los datos obtenidos. Las series seleccionadas corresponden a los precios de cierre de las acciones de Apple (APPL), Amazon (AMZN) y al índice bursátil (SP). Estas series ofrecen una buena muestra de cómo los modelos pueden desempeñarse en distintos contextos de mercado, desde gigantes tecnológicos hasta un índice regional.

Utilizamos varias métricas para evaluar el rendimiento de cada modelo, como el Error Cuadrático Medio (MSE), el Error Absoluto Medio (MAE) y el Coeficiente de Determinación (R^2). Note a continuación un análisis detallado de los resultados obtenidos, resaltando las fortalezas y debilidades de los modelos bajo cada uno de los espacios.

1. ANÁLISIS DE LOS RESULTADOS OBTENIDOS PARA LA SERIE APPLE

A partir de los resultados obtenidos para el mejor modelo y kernel evaluado por el desempeño obtenido al realizar la aplicación de estos, se puede realizar un análisis detallado para cada uno de los indicadores de evaluación de efectividad y rendimiento de cara a la predicción sobre la volatilidad en la serie de tiempo tanto en su versión original como la estimación realizada en el espacio de Hilbert, recuerde que esta serie es particular por los picos de volatilidad que presenta aun sin tener una masa de datos muy grande respecto a las demás analizadas, una vez claro lo anterior, los resultados de los indicadores aplicados a los modelos fueron los siguientes:

Tabla 9 Resultados evaluación del modelo para la serie APPL

Método	**MSE**	**MAE**	$\boldsymbol{R^2}$
AR con EWMA (Original)	0.000066	12.516495	0.584264
AR con EWMA (Hilbert)	7.785708	1.303766	0.994486

Basado en la tabla 8 se puede inferir que, el modelo AR con EWMA en el espacio de Hilbert muestra un MSE significativamente mayor en comparación con el modelo original. Esto sugiere que el modelo en el espacio de Hilbert está ajustando los datos con menor precisión en términos de MSE, para este caso puntual es importante hacer hincapié sobre la masa que presenta la data la cual no es amplia para una mayor adaptabilidad del modelo en este espacio, si bien se puede percibir como una debilidad en la aplicación, note que el indicador de desempeño MAE es considerablemente menor en el espacio de Hilbert comparado con el modelo original, indicando que, lo que permite afirmar que en promedio, las predicciones del modelo en Hilbert están más cerca de los valores reales lo que hace que la predicción de la volatilidad no provea un margen alto de error y esto se puede asegurar al observar que, el valor de R^2 para el modelo en el espacio de Hilbert es muy alto, lo que sugiere que este modelo bajo el espacio de Hilbert explica, en su mayoría, toda la variabilidad en los datos.

Durante el análisis de la serie en términos computacionales, el proceso de evaluación de costo computacional se llevó a cabo en dos etapas, una dentro y otra fuera del espacio de Hilbert para cada aplicación de los modelos AR. Dentro del espacio de Hilbert, se evidenció que constó de 11 iteraciones, tomó aproximadamente 120 segundos en completarse. Por otra parte, sin el mapeo al espacio de Hilbert, la aplicación del modelo contó con 20 iteraciones, requirió aproximadamente de 240 segundos. La eficiencia del costo computacional se optimizó significativamente, logrando convergencia de los parámetros en un tiempo razonable evitando un consumo excesivo de recursos [48].

Así entonces por los anteriores indicadores analizados se puede afirmar que, aunque el MSE es mayor en el espacio de Hilbert, el MAE y el R^2 indican que el modelo AR con EWMA en este espacio proporciona predicciones más cercanas y explica la variabilidad de los datos con una alta precisión. De tal modo es posible concluir que la transformación al espacio de Hilbert ha mejorado la capacidad del modelo para capturar la volatilidad en la serie de tiempo; proviniendo del modelo con mejor desempeño para esta serie previamente, la aplicación al espacio de Hilbert otorga un sobreajuste para elevar la efectividad resultante.

2. ANÁLISIS DE LOS RESULTADOS OBTENIDOS PARA LA SERIE AMAZON

Para la serie Amazon, la tabla 9 muestra un contraste interesante entre el modelo AR con EWMA tradicional y su versión en el espacio de Hilbert utilizando un kernel RBF. El modelo original, sin la transformación a un espacio de Hilbert, presenta un MSE (error cuadrático medio) de 4.952651e-07 y un MAE (error absoluto medio) de 7.512885e+00, con un coeficiente de determinación R^2 de 0.923388. Por otro lado, el modelo en el espacio de Hilbert arroja un MSE de 2.425973e-32 y un MAE de 2.985789e-13, con un R^2 de 1.000000.

Tabla 10 Resultados evaluación del modelo para la serie AMZN

Método	**MSE**	**MAE**	$\boldsymbol{R^2}$
AR con EWMA (Original)	4.952651e-07	7.512885	0.923388
AR con EWMA (Hilbert)	425973e-32	2.985789e-13	1.000000

Estos resultados sugieren que la transformación a un espacio de Hilbert, utilizando el kernel RBF, ha mejorado significativamente el ajuste del modelo a los datos. El MSE y el MAE han disminuido drásticamente, lo que indica que las predicciones del modelo en el espacio de Hilbert son mucho más precisas y cercanas a los valores reales. Además, el valor de R^2 cercano a 1 indica que el modelo explica prácticamente toda la variabilidad de los datos, lo que sugiere un ajuste casi perfecto.

En el análisis de la serie se llevó a cabo un proceso de evaluación del costo computacional en dos fases, utilizando diferentes modelos AR fuera y dentro del espacio de Hilbert. La primera fase alcanzó un valor de función de -345.40196688514027 tras 19 iteraciones, con 133 evaluaciones de la función y 15 del gradiente, completándose en aproximadamente 390 segundos. En la

segunda fase, con el mapeo al espacio de Hilbert, se logró un valor de función de -208.026308090331 luego de 13 iteraciones, con 80 evaluaciones de la función y 12 del gradiente, requiriendo aproximadamente 240 segundos. La eficiencia del costo computacional se mejoró notablemente, logrando la convergencia de los parámetros en un tiempo razonable sin necesidad de un consumo excesivo de recursos [48].

La superioridad del modelo en el espacio de Hilbert puede atribuirse a varias razones. El kernel RBF permite capturar relaciones no lineales y patrones complejos en los datos que podrían no ser evidentes en el espacio original. Al proyectar los datos en un espacio de mayor dimensión, el modelo puede encontrar representaciones más adecuadas para capturar la dinámica de la volatilidad de los precios de Amazon. Además, la transformación a un espacio de Hilbert puede ayudar a mitigar el efecto de ruido y outliers en los datos, mejorando así la precisión de las predicciones.

3. ANÁLISIS DE LOS RESULTADOS OBTENIDOS PARA LA SERIE BVE (SP)

Es importante iniciar este análisis de resultados recordando que el modelo que presentó mejor desempeño para el índice SP de la serie trabajada sobre la BVE fue el AR bajo la técnica de suavizamiento EWMA, este a su vez se trabajó bajo el kernel RBF que fue el mejor ubicado frente a los demás kernels evaluados, note lo anterior en los capítulos **VI** y **VII,** ahora teniendo claras las condiciones del modelo aplicado al espacio de Hilbert con Kernel reproductivo es valioso aclarar el contexto de la serie temporal principalmente el índice SP el cual fue tomado para realizar todo el bosquejo trabajado, si bien esta serie de tiempo cuenta con una volatilidad alta, semejante a sus iguales evaluadas anteriormente, como lo fueron las series de Apple y Amazon, esta serie cuanta con un punto importante a destacar y es su alta masa de información debido a que es un índice bursátil internacional no solo de inversión sino también de análisis de mercado, lo cual cambia el paradigma de análisis de los resultados, esto pues una alta masa de información en una serie de tiempo se traduce en una mayor adaptabilidad de los modelos bajo el espacio de Hilbert [22]. Lo anterior permite afirmar que esta serie prima en el análisis de sus resultados como el pilar para este estudio, debido a que por su alta capacidad de adaptabilidad que ofrece a los modelos se puede plasmar con mayor claridad los resultados obtenidos, así entonces los indicadores de desempeño de los modelos que se obtuvieron para esta serie bajo el espacio de Hilbert fueron los siguientes:

Tabla 11 Resultados evaluación del modelo para la serie BVE-SP

Método	MSE	MAE	R^2
AR con EWMA (Original)	4.952651e-07	7.715747e-03	-0.0190351
AR con EWMA (Hilbert)	2.425973e-32	1.999071e-16	1.000000

Al observar la tabla 10 se puede afirmar que, el modelo AR con suavizamiento EWMA en el espacio de Hilbert tiene un MSE extremadamente bajo comparado con el modelo original. Esto sugiere a primera vista que la transformación al espacio de Hilbert

permite capturar mejor las características de la serie de tiempo, lo cual era un escenario esperado por lo aclarado anteriormente, esto permite lograr predicciones muy precisas. Ahora note que, el indicador MAE también es significativamente menor en el espacio de Hilbert en comparación con el modelo original este resultado refuerza la precisado anteriormente acerca de la efectividad de las predicciones en el espacio de Hilbert, finalmente fíjese que el valor del indicador R^2 para el modelo en el espacio de Hilbert es perfecto, entiéndase perfecto por la unidad precisa alcanzada en su resultado, indicando un ajuste exacto a los datos. En contraste, el modelo original tiene un valor de el R^2 negativo, lo que sugiere que el modelo sin su aplicación al espacio de Hilbert no logra capturar la variabilidad de los datos adecuadamente.

Al centrar el análisis en la eficiencia del costo computacional se obtuvo que, sin el mapeo al espacio de Hilbert, se alcanzó un valor de función de -912.9659901240254 después de 6 iteraciones, con 69 evaluaciones de la función y 5 del gradiente, completándose en aproximadamente 980 segundos. Con el mapeo al espacio de Hilbert, se logró un valor de función de -1334.8486595842705 luego de 14 iteraciones, con 114 evaluaciones de la función y 10 del gradiente, tomando aproximadamente 150 segundos. Además, una prueba adicional con mapeo al espacio de Hilbert arrojó un valor de -1468.5463290191137 en tan solo 1 iteración, con 5 evaluaciones de la función y 1 del gradiente, completándose en aproximadamente 10 segundos. Estos resultados destacan una mejora significativa en la eficiencia del costo computacional [48].

X. CONCLUSIONES Y TRABAJOS FUTUROS

A. CONCLUSIONES

En este trabajo se ha explorado una metodología específica para el preprocesamiento y acondicionamiento de bases de datos de series temporales financieras esto en busca de lograr el objetivo principal que radica en optimizar la calidad y relevancia de los datos financieros, preparándolos de manera efectiva para su posterior análisis en el contexto de proyectos financieros. Note el capítulo IV; al realizar el preprocesamiento fue posible identificar características esenciales de las series de tiempo que se trabajarían lo que permitió darle un contexto a los futuros análisis que se desarrollarían en estas mismas, a partir de este hilo conductor es posible concluir que gracias a la metodología explorada los resultados obtenidos en los capítulos siguientes cobraron mayor sentido y relevancia, esto da paso a poder afirmar que la limpieza de los datos fue un factor esencial para obtener una base de datos confiable, a su vez este proceso implicó manejar valores atípicos, tratar datos faltantes y corregir errores de medición. La transformación de datos fue adapta a la información de las necesidades del análisis, incluyendo normalización, estandarización y diferenciación. Además, se tuvo que considerar la estacionalidad inherente a las series de tiempo financieras analizadas y aplicar métodos para desestacionalizar o modelar estas mismas.

Sobre los capítulos **VI** a **VIII** recae la responsabilidad de poder brindar un mensaje claro a la necesidad de construir un esquema de estimación de la volatilidad financiera lo suficientemente robusto que permitiese brindar la representación de modelos autorregresivos con múltiples opciones a evaluar para poder reproducir estos mismos en espacios de Hilbert con un kernel que ofreciera un modelado efectivo, para el desarrollo de este imperativo del trabajo se contó con distintos modelos que fueron evaluados a cada una de las series, note que gracias a los modelos autorregresivos observados en el capítulo **VI** fue posible identificar dos hitos importantes, el primero fue la contextualización de cada ejercicio respecto a la serie de tiempo a tratar, si bien, las series de Apple y Amazon contaban con características particulares, estas dos a su vez reflejaban en esencia un índice financiero con una volatilidad alta pero con una pendiente constante en el crecimiento exponencial de los valores históricos, ahora en el caso de la serie de tiempo BVE para el índice SP se contaba con una particularidad, la volatilidad reflejaba comportamientos similares a las demás series pero el tamaño de la información que otorgaba esta serie era significativamente mayor al de las otras series por lo que la posicionaba como un posible pilar de análisis que permitiría dar un mensaje claro a la necesidad trabajada. De este modo la identificación de un modelo que se adaptara a estas series se podría lograr al evaluar el indicador MSE; lo que permite dar paso al segundo hito importante, si bien, ya se contaba con un contexto a cada modelo realizado en las series de tiempo fue posible identificar que, para las series de tiempo observadas el mejor desempeño se obtuvo al aplicar un modelo AR no obstante el modelo sin suavizamiento no fue del todo preciso por lo que gracias a estos comportamientos observados es posible concluir que los modelos que mejor adaptabilidad presenten, como lo fueron los modelos AR y ARCH, se deben trabajar con suavizamiento EWMA lo que permite reducir la brecha diferencial entre el estimado de volatilidad y la recta real de la serie. Una vez claro el modelo con mejor rendimiento, gracias al capítulo **VII** fue posible inferir que el aplicar un kernel reproductivo

ha demostrado ser altamente efectivo para la estimación de la volatilidad financiera gracias a la métrica MSE, esto permite que se optimice el grado de asertividad del modelo por lo que al realizar distintas pruebas se pudo evaluar que los kernel reproductivos con mayor rendimiento fueron los kernel polinomial y RBF con un ajuste óptimo para las necesidades de cada serie.

En consecuencia de lo anterior para poder reproducir estos mismos pasos en espacios de Hilbert con un kernel que ofreciera un modelado efectivo se utilizaron los resultados de los capítulos **VI** y **VII** para trabajar con el modelo y kernel mejor ubicado en términos de efectividad y rendimiento respectivamente, por lo que para el capítulo **VIII** se trabajó con un único modelo para cada serie el cual, gracias a lo mencionado anteriormente, contaba con una adaptación previa a la misma; de este modo al obtener los resultados bajo el espacio de Hilbert fue posible observar que, aunque la precisión de los modelos ya era alta; en el espacio de Hilbert cada uno de los kernel presentó un mejor rendimiento lo que permitió alcanzar una precisión significativamente superior respecto con la que se contaba previamente.

Respecto a lo mencionado previamente, es posible afirmar que un esquema de estimación de la volatilidad financiera ideal que permita brindar la representación de los modelos autorregresivos debe contar con una preselección bajo rendimiento y calidad de efectividad para cada uno de los modelos, no obstante, algunos de los mejores rendimientos alcanzados fueron evidenciados en modelos tales como AR y ARCH bajo suavizamiento EWMA. Adicional a esto, es posible indicar que una vez se cuente con el mejor modelo para cada serie que se esté analizando se puede obtener un mejor rendimiento del kernel y así lograr una mayor precisión en la estimación gracias a la reproducción de estos mismos en espacios de Hilbert.

Una vez claro el cómo los modelos autorregresivos pueden ser optimizados mediante la aplicación de espacios de Hilbert para una mejor predicción de la volatilidad en una serie de tiempo financiera, toma suma importancia el saber cuáles serían las medidas de desempeño ideales que permitan evaluar la confiabilidad del modelo desarrollado, para esto note el capítulo **IX** donde además del MSE como indicador calificativo de los resultados obtenidos se trabajó con dos indicadores más como lo fueron el MAE y R^2 los cuales permitían obtener una visión más amplia a los resultados obtenidos y así determinar su efectividad, es importante recordar que, la elección del modelo y su evaluación debían estar alineadas con los objetivos específicos del análisis y las necesidades del mercado o el contexto de los datos trabajados, de este modo el indicador MAE o Error Absoluto Medio también, al ser una métrica fácil de interpretar porque utiliza la misma escala de los datos evaluados permitió en cada uno de los ejercicios identificar cual era la magnitud de cada error, por otro lado el R^2 o coeficiente de determinación al indicar la proporción de variabilidad de la variable dependiente para cada modelo permitía calificar la asertividad de los resultados, por lo que la combinación de estas dos métricas junto con el MSE se pueden asumir como las medidas de desempeño ideales que permiten evaluar la confiabilidad del modelo.

Finalmente, como resultado de las observaciones anteriores, este trabajo ha demostrado la eficacia de utilizar modelos autorregresivos y espacios de Hilbert para estimar la volatilidad y el riesgo de mercado en series financieras. Al implementar

herramientas basadas en ciencia de datos, se ha logrado optimizar la calidad de los datos y, a su vez, se han aplicado técnicas avanzadas como lo fue el suavizamiento EWMA para mejorar la precisión de los modelos. Los resultados al contar con una evaluación implementada bajo métricas que validan la precisión, desempeño y rendimiento muestran que el uso de kernels en los modelos autorregresivos en espacios de Hilbert, ofrece un modelado efectivo y mejora significativamente la precisión de las estimaciones.

B. TRABAJOS FUTUROS

1. **Diversificación de Datos:** Extender la metodología a diferentes tipos de series temporales financieras, como lo son las criptomonedas.
2. **Optimización de Kernels:** Investigar y desarrollar nuevos kernels para espacios de Hilbert que ofrezcan un mejor rendimiento y precisión.
3. **Integración con Machine Learning:** Implementar técnicas avanzadas de machine learning y deep learning para mejorar la predicción y modelado de series temporales financieras.
4. **Evaluación de Riesgo en Tiempo Real:** Desarrollar sistemas en tiempo real para la estimación de volatilidad y riesgo, utilizando la metodología propuesta.
5. **Automatización del Preprocesamiento de Datos:** Crear herramientas automatizadas para el preprocesamiento de grandes volúmenes de datos financieros.
6. **Análisis Comparativo:** Realizar análisis comparativos entre diferentes metodologías de suavizamiento para determinar cuáles ofrecen mejores resultados.

XI. REFERENCIAS BIBLIOGRÁFICAS

[1] C. A. Soto Campos, V. A. Reyes Rodríguez, y C. Rondero Guerrero, “Series de tiempo financieras. Un enfoque alternativo”, *Pädi Boletín Científico de Ciencias Básicas e Ingenierías del ICBI*, vol. 7, núm. 14, 2020, doi: 10.29057/icbi.v7i14.4430.

[2] T. Bollerslev, “Reprint of: Generalized Autoregressive Conditional Heteroskedasticity”, *J Econom*, vol. 234, 2023, doi: 10.1016/j.jeconom.2023.02.001.

[3] E. R. Ziegel y W. Enders, “Applied Econometric Time Series”, *Technometrics*, vol. 37, núm. 4, 1995, doi: 10.2307/1269759.

[4] R. S. Tsay, *Analysis of financial time series*. 2010. doi: 10.1002/9780470644560.

[5] R. Montenegro, “Medición de la volatilidad en series de tiempo financieras: Una evaluación a la tasa de cambio representativa del mercado (TRM) en Colombia”, *Finanazas y Política Económica*, vol. 2, núm. 1, 2010.

[6] G. Hyndman, R.J., & Athanasopoulos, “Forecasting: Principles and Practice (3rd ed)”, *OTexts: Melbourne, Australia. OTexts.com/fpp3*, 2015.

[7] N. G. Mankiw, *Microeconomics 7th Edition*. 2013.

[8] B. Schölkopf y A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond Adaptive computation and machine learning*. 2001.

[9] R. Y. Goh, G. Andreeva, y Y. Cao, “Predicting Financial Volatility from Personal Transactional Data”, en *2022 IEEE Symposium on Computational Intelligence for Financial Engineering and Economics, CIFEr 2022 - Proceedings*, 2022. doi: 10.1109/CIFEr52523.2022.9776206.

[10] J. M. Keynes, *The general theory of employment, interest, and money*. 2018. doi: 10.1007/978-3-319-70344-2.

[11] D. Kú, M. Trigueros, y A. Oktaç, “Comprensión del concepto de base de un espacio vectorial desde el punto de vista de la teoría APOE CORE”, *Funes*, vol. 20, núm. 2, 2008.

[12] R. Bruzual y M. Domínguez, “Problemario Espacios de Hilbert”, Caracas, 2005.

[13] Y. Goldberg, “A primer on neural network models for natural language processing”, *Journal of Artificial Intelligence Research*, vol. 57, 2016, doi: 10.1613/jair.4992.

[14] R. A. Purba, S. Samsir, M. Siddik, S. Sondang, y M. F. Nasir, “The optimalization of backpropagation neural networks to simplify decision making”, en *IOP Conference Series: Materials Science and Engineering*, 2020. doi: 10.1088/1757-899X/830/2/022091.

[15] J. García, J. Jiménez, y J. Pérez, “Predicción de series financieras mediante bosques aleatorios”, *Revista Economía Aplicada*, vol. 29, pp. 299–324, 2021.

[16] V. M. Ortiz Rosas, “Estimación del comportamiento de la volatilidad a través de series de tiempo y medidas de riesgo”, *BUAP - Benemérita Universidad Autónoma de Puebla*, oct. 2022, doi: https://hdl.handle.net/20.500.12371/16992.

[17] S. Velez Hernandez, “Modelos híbridos para el pronostico de volatilidad y su uso en medidas de riesgo”, 2023. Consultado: el 18 de noviembre de 2023. [En línea]. Disponible en: https://repository.urosario.edu.co/server/api/core/bitstreams/3705ab3a-10b9-4d99-b15e-25c428e18af0/content

[18] A. Ferrando Esteve, “Predicción del comportamiento de criptomonedas a partir de métodos de inteligencia artificial”, sep. 2023, Consultado: el 19 de noviembre de 2023. [En línea]. Disponible en: https://riunet.upv.es:443/handle/10251/197268

[19] M. del C. Franco Rodríguez y E. Velasteguí López, “El suavizado exponencial en el pronóstico de series no estacionarias.”, *Ciencia Digital*, vol. 4, núm. 1, 2020, doi: 10.33262/cienciadigital.v4i1.1073.

[20] R. Demirer y K. Gkillas, “Risk aversion and the predictability of crude oil market volatility: A forecasting experiment with random forests”, *Journal of the Operational Research Society* , vol. 73, 2021.

[21] G. López Pineda, “Modelos de regresión para datos funcionales por la metodología de Kernel reproductor en espacios de Hilbert”, *BUAP*, 2017, Consultado: el 7 de diciembre de 2023. [En línea]. Disponible en: https://hdl.handle.net/20.500.12371/488

[22] E. A. Valencia Angulo, *Métricas entre procesos aleatorios usando el método de embebimiento de distribuciones de probabilidad en un espacio de Hilbert con kernel reproductivo*. 2022. doi: 10.22517/9789587227796.

[23] P. Sang y B. Li, “Nonlinear function-on-function regression by RKHS”, *Cornell University*, jul. 2022, [En línea]. Disponible en: http://arxiv.org/abs/2207.08211

[24] M. F. Villa Gil, “Estimación y predicción del modelo ARCH para la volatilidad a través de embebimientos de distribuciones de probabilidad en espacios de Hilbert con kernel reproductivo”, *Universidad Tecnologica de Pereira*,

2022.
[25] C. Bergmeir, R. J. Hyndman, y J. M. Benítez, “Bagging exponential smoothing methods using STL decomposition and Box-Cox transformation”, *Int J Forecast*, vol. 32, núm. 2, 2016, doi: 10.1016/j.ijforecast.2015.07.002.
[26] V. Kuznetsov y M. Mohri, “Discrepancy-Based Theory and Algorithms for Forecasting Non-Stationary Time Series”, *Ann Math Artif Intell*, vol. 88, núm. 4, 2020, doi: 10.1007/s10472-019-09683-1.
[27] T. Hofmann, B. Schölkopf, y A. J. Smola, “Kernel methods in machine learning”, 2008. doi: 10.1214/009053607000000677.
[28] H. N. Mhaskar y T. Poggio, “Deep vs. shallow networks: An approximation theory perspective”, *Analysis and Applications*, vol. 14, núm. 6, 2016, doi: 10.1142/S0219530516400042.
[29] O. Akbilgic, “ISTANBUL STOCK EXCHANGE”, UC Irvine Machine Learning Repository.
[30] P. Mooney, “Stock Market Data (NASDAQ, NYSE, S&P500)”, Kaggle.
[31] A. Amato y V. Di Lecce, “Data preprocessing impact on machine learning algorithm performance”, *Open Computer Science*, vol. 13, núm. 1, 2023, doi: 10.1515/comp-2022-0278.
[32] M. Geurts, G. E. P. Box, y G. M. Jenkins, “Time Series Analysis: Forecasting and Control”, *Journal of Marketing Research*, vol. 14, núm. 2, 1977, doi: 10.2307/3150485.
[33] L. Seymour, P. J. Brockwell, y R. A. Davis, “Introduction to Time Series and Forecasting.”, *J Am Stat Assoc*, vol. 92, núm. 440, 1997, doi: 10.2307/2965440.
[34] R. F. Engle, “Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation”, *Econometrica*, vol. 50, núm. 4, 1982, doi: 10.2307/1912773.
[35] R. Engle, “Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models”, *Journal of Business and Economic Statistics*, vol. 20, núm. 3, 2002, doi: 10.1198/073500102288618487.
[36] E. S. Gardner, “Exponential smoothing: The state of the art-Part II”, *Int J Forecast*, vol. 22, núm. 4, 2006, doi: 10.1016/j.ijforecast.2006.03.005.
[37] S. Russell y P. Norvig, *Artificial Intelligence, Global Edition A Modern Approach*. 2021.
[38] F. Black, “Studies in stock price volatility changes”, 1976.
[39] Books, *Introductory to functional analysis with applications (Kreyszig)*. 1987.
[40] C. E. Rasmussen, “Gaussian Processes in machine learning”, *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 3176, 2004, doi: 10.1007/978-3-540-28650-9_4.
[41] N. Cristianini y J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. 2000. doi: 10.1017/cbo9780511801389.
[42] F. Riesz, “Sur la théorie ergodique”, *Commentarii Mathematici Helvetici*, vol. 17, núm. 1, 1944, doi: 10.1007/BF02566244.
[43] T. W. Gamelin y J. B. Conway, “A Course in Functional Analysis.”, *The American Mathematical Monthly*, vol. 93, núm. 9, 1986, doi: 10.2307/2322311.
[44] E. Parzen, “An Approach to Time Series Analysis”, *The Annals of Mathematical Statistics*, vol. 32, núm. 4, 1961, doi: 10.1214/aoms/1177704840.
[45] J. Shawe-Taylor y N. Cristianini, *Kernel Methods for Pattern Analysis*. 2004. doi: 10.1017/cbo9780511809682.
[46] C. M. Bishop, “Pattern Recognition and Machine Learning: 06 - Kernel Methods”, *Pattern Recognition and Machine Learning*, 2006.
[47] T. Hastie, R. Tibshirani, y J. Friedman, *Springer Series in Statistics The Elements of Statistical Learning - Data Mining, Inference, and Prediction*, vol. 2nd. 2009.
[48] D. Cárdenas y G. Quintero, “GitHub Repository”, GitHub Repository.
[49] J. C. Alvarado-Pérez, D. H. Peluffo-Ordóñez, y R. Theron, “Visualización y métodos kernel: Integrando inteligencia natural y artificial”, *N/a*, 2016.
[50] S. R. Sain y V. N. Vapnik, “The Nature of Statistical Learning Theory”, *Technometrics*, vol. 38, núm. 4, 1996, doi: 10.2307/1271324.
[51] H. Akaike, “Information Theory and an Extension of the Maximum Likelihood Principle”, 2015. doi: 10.1007/978-1-4612-0919-5_38.
[52] R. E. Kalman, “A new approach to linear filtering and prediction problems”, *Journal of Fluids Engineering, Transactions of the ASME*, vol. 82, núm. 1, 1960, doi: 10.1115/1.3662552.
[53] D. B. Nelson, “Conditional Heteroskedasticity in Asset Returns: A New Approach”, *Econometrica*, vol. 59, núm. 2, 1991, doi: 10.2307/2938260.

[54] R. T. Baillie y T. Bollerslev, “The message in daily exchange rates: A conditional-variance tale”, *Journal of Business and Economic Statistics*, vol. 7, núm. 3, 1989, doi: 10.1080/07350015.1989.10509739.