

Mejora de la predicción de precios en criptomonedas mediante redes neuronales transformadoras e indicadores técnicos

Mohammad Ali Labbaf Khaniki*, Mohammad Manthouri

Facultad de Ingeniería Eléctrica, Universidad Tecnológica K.N. Toosi, Teherán, Irán

*mohammadlabbaf@email.kntu.ac.ir

Resumen: Este estudio presenta un enfoque innovador para predecir series temporales de criptomonedas, centrándose específicamente en Bitcoin, Ethereum y Litecoin. La metodología integra el uso de indicadores técnicos, una red neuronal Performer y BiLSTM (Bidirectional Long Short-Term Memory) para captar la dinámica temporal y extraer características significativas de los datos brutos de criptomonedas. La aplicación de indicadores técnicos facilita la extracción de patrones intrincados, impulso, volatilidad y tendencias. La red neuronal Performer, que emplea la atención rápida mediante características aleatorias ortogonales positivas (FAVOR+), ha demostrado una eficiencia computacional y una escalabilidad superiores en comparación con el mecanismo tradicional de atención multicabezal en los modelos Transformer. Además, la integración de BiLSTM en la red feedforward mejora la capacidad del modelo para captar la dinámica temporal de los datos, procesándolos tanto hacia delante como hacia atrás. Esto resulta especialmente ventajoso para datos de series temporales en los que los puntos de datos pasados y futuros pueden influir en el estado actual. El método propuesto se ha aplicado a los plazos horario y diario de las principales criptomonedas y su rendimiento se ha comparado con el de otros métodos documentados en la bibliografía. Los resultados subrayan el potencial del método propuesto para superar a los modelos existentes, lo que supone un avance significativo en el campo de la predicción del precio de las criptomonedas.

Palabras clave: Criptomoneda, Deep Learning, Predicción de series temporales, Transformador, Ejecutor, Mecanismo de atención,

1) Introducción

En el panorama en rápida evolución de la tecnología, el modo de las transacciones ha experimentado un importante cambio de paradigma. Los pagos físicos tradicionales, como el efectivo y los cheques, están siendo sustituidos cada vez más por transacciones digitales. Esta transformación se ha visto impulsada en gran medida por la llegada y proliferación de las criptomonedas, que han surgido como una nueva clase de activos y medio de intercambio (Aghashahi y Bamdad, 2023). Las criptomonedas, a diferencia de las monedas fiduciarias convencionales, emplean claves criptográficas para facilitar las transacciones financieras. En la última década, las finanzas digitales han experimentado un crecimiento exponencial, y las criptomonedas han liderado este avance innovador. Uno de los aspectos más críticos del uso de cualquier moneda, ya sea como medio de transacción o como activo, es la capacidad de predecir su valor esperado. El valor y la estabilidad de cualquier moneda dependen en gran medida de la autoridad que la controla. En el caso de las monedas fiduciarias, suele ser el gobierno del país correspondiente. Sin embargo, las criptomonedas operan de forma descentralizada, libres del control gubernamental. Esta característica única presenta tanto oportunidades como retos. Por un lado, ofrece un gran potencial de rentabilidad y diversificación. Esto subraya la necesidad de modelos de predicción de precios sólidos y precisos en el ámbito de las criptodivisas (Pichaiyuth *et al.*, 2023).

Predecir el precio de las criptomonedas, como Bitcoin, presenta una serie de retos únicos. La principal dificultad reside en la volatilidad inherente al mercado de las criptomonedas. A diferencia de los mercados financieros tradicionales, el mercado de criptomonedas está abierto 24 horas al día, 7 días a la semana, lo que provoca cambios de precios más frecuentes. Además, el precio de las criptomonedas puede verse influido por diversos factores, como los avances tecnológicos, las noticias sobre regulación, la confianza del mercado y las tendencias macroeconómicas (Zhao, Crane y Bezbradica, 2022). Además, la falta de una autoridad

centralizada añade otro nivel de complejidad, ya que el valor no está vinculado a un activo físico ni controlado por una única autoridad.

entidad. Estos factores contribuyen a la imprevisibilidad de los precios de las criptomonedas, lo que hace que una predicción precisa sea una tarea compleja. A pesar de estos retos, el enorme valor potencial de las criptomonedas ha atraído una gran atención tanto de inversores como de investigadores, lo que ha llevado al desarrollo de sofisticados modelos de predicción (Awoke *et al.*, 2021).

Los indicadores técnicos desempeñan un papel crucial en el comercio y el análisis de criptodivisas. Proporcionan a los operadores un enfoque estadístico para evaluar las condiciones del mercado y predecir las tendencias de los precios. Indicadores como el Índice de Fuerza Relativa (RSI) y las medias móviles se utilizan comúnmente para identificar posibles señales de compra o venta. Por ejemplo, el RSI puede ayudar a determinar si una criptomoneda está sobrecomprada o sobrevendida, mientras que las medias móviles pueden señalar el inicio de una tendencia alcista o bajista. Además, estos indicadores pueden utilizarse junto con modelos de aprendizaje automático para mejorar la precisión de las predicciones (Pichaiyuth *et al.*, 2023). Al extraer características significativas de los datos de precios en bruto, permiten a estos modelos captar patrones y tendencias complejos, mejorando así la eficacia de la predicción de los precios de las criptomonedas. Estos indicadores técnicos pueden servir como valiosas aportaciones a los modelos de predicción. Al captar propiedades estadísticas clave del mercado, pueden ayudar a los modelos a identificar posibles movimientos futuros en los precios de las criptomonedas (Goutte *et al.*, 2023).

Los transformadores, una arquitectura de red neuronal, han despertado un interés considerable tanto en el procesamiento del lenguaje natural (PLN) como en el análisis de series temporales. Su capacidad para gestionar dependencias de largo alcance y el procesamiento paralelo han impulsado su popularidad en estos ámbitos (Rahali y Akhloufi, 2023). Utilizando la autoatención multicabezal o la atención de producto de punto escalado, los Transformers pueden evaluar la importancia de las entradas dentro de una secuencia, lo que permite la captura de

patrones de datos intrincados. En PLN, los transformadores han sido fundamentales para lograr un rendimiento de vanguardia en tareas como la traducción, el resumen y el análisis de sentimientos (Patwardhan, Marrone y Sansone, 2023). En la predicción de series temporales, los Transformers son prometedores, ya que aprovechan su capacidad para captar

relaciones temporales. Sus aplicaciones abarcan desde la previsión meteorológica y del mercado de criptomonedas hasta la detección de fallos y anomalías. Puede identificar patrones y tendencias a lo largo del tiempo, lo que permite predecir puntos de datos futuros con un alto grado de precisión (Haryono, Sarno y Sungkono, 2023).

El mecanismo de atención desempeña un papel fundamental en la arquitectura de Transformer, mejorando significativamente su rendimiento en la predicción de series temporales (Vaswani *et al.*, 2017). El mecanismo de atención permite que el modelo se centre en diferentes partes de la secuencia de entrada al producir una salida, capturando eficazmente las dependencias entre palabras o eventos que están muy alejados. En la predicción de series temporales, el mecanismo de atención permite al Transformer sopesar la importancia de los eventos pasados a la hora de predecir los futuros. Esto resulta especialmente útil en situaciones en las que los acontecimientos recientes pueden no ser los más relevantes para realizar una predicción. Además, el mecanismo de atención de los Transformers es eficiente desde el punto de vista computacional, ya que permite el cálculo paralelo a lo largo de la secuencia, a diferencia de las RNN (Redes Neuronales Recurrentes), que requieren un cálculo secuencial (Samii *et al.*, 2023). Esto hace que Transformers sea más rápido y escalable para grandes conjuntos de datos (Labba Khaniki, Mirzaeibonehkhater y Manthouri, 2023).

Esta investigación presenta una metodología pionera para la predicción de series temporales de Bitcoin, Ethereum y Litecoin. El enfoque utiliza inicialmente indicadores técnicos para extraer características estadísticas de los datos. A continuación, se aplica una red neuronal Performer, que utiliza atención rápida mediante características aleatorias ortogonales positivas (FAVOR+) en lugar del mecanismo de atención multicabezal utilizado en Transformer. La arquitectura incluye además BiLSTM, lo que mejora su capacidad para captar la dinámica temporal. Los principales avances de esta investigación se explicarán en las secciones siguientes. A continuación se describen los principales avances de este método:

1. **Extracción de características mediante indicadores técnicos:** En el ámbito del análisis financiero, los indicadores técnicos como el RSI y las medias móviles se utilizan a menudo para predecir los movimientos futuros de los precios basándose en datos históricos. Mediante el uso de estos indicadores, la investigación extrae características significativas de los datos brutos de criptomonedas, que pueden captar patrones y tendencias complejos que podrían pasar desapercibidos a simple vista.
2. **Performer:** En esta investigación se da un paso más incorporando una red neuronal Performer. El mecanismo FAVOR+ es más eficiente computacionalmente y escalable que el mecanismo de atención multicabezal. Esto se debe a que aproxima el mecanismo de atención de una forma que requiere menos recursos computacionales, lo que permite procesar secuencias de datos más grandes. Esto es especialmente beneficioso en tareas como la predicción de series temporales, en las que el modelo necesita procesar largas secuencias de datos históricos para hacer predicciones precisas.
3. **BiLSTM:** el uso de BiLSTM (Bidirectional Long Short-Term Memory, memoria bidireccional a corto plazo) en la red feedforward permite a su modelo capturar la dinámica temporal de los datos, ya que puede procesar los datos en ambas direcciones, hacia delante y hacia atrás. Esto es especialmente útil para los datos de series temporales, en los que los puntos de datos pasados y futuros pueden influir en el estado actual. Por otro lado, las capas totalmente conectadas permiten al modelo aprender complejas relaciones no lineales entre las características. Esta combinación de BiLSTM y capas totalmente conectadas hace que su modelo sea capaz de manejar la complejidad y la volatilidad que a menudo se observa en los movimientos de los precios de las criptomonedas.

Los novedosos componentes de esta investigación forman colectivamente un método avanzado para predecir series temporales de criptomonedas, superando potencialmente el

rendimiento predictivo de los modelos existentes. Para evaluar nuestro enfoque, lo aplicamos a los marcos temporales diarios y horarios de las principales criptomonedas, como Bitcoin, Ethereum y Litecoin. A continuación, comparamos su rendimiento con

otros métodos documentados en la literatura, específicamente los presentados en (Awoke *et al.*, 2021), (Jay *et al.*, 2020), y (Girsang y Stanley, 2023).

La organización de este documento es la siguiente: La Sección II ofrece una revisión de la literatura relevante. La Sección III profundiza en la metodología propuesta, analizando el uso de indicadores financieros para la extracción de características y el método de aprendizaje profundo propuesto que combina BiLSTM y Performer. La Sección IV muestra los resultados de la simulación, proporcionando detalles sobre el proceso de entrenamiento y evaluación del enfoque propuesto. Por último, la Sección V resume las principales conclusiones y contribuciones de nuestro estudio.

2) Obras relacionadas

En las últimas décadas, el aprendizaje automático y el aprendizaje profundo han provocado cambios sustanciales en el campo de las previsiones financieras, incluida la predicción de los precios de las criptomonedas (Zhang *et al.*, 2021) y (Mohammadabadi *et al.*, 2023). Los modelos de aprendizaje automático utilizan datos de precios históricos para anticipar tendencias futuras, capturando eficazmente patrones intrincados y no lineales en los datos. El aprendizaje profundo, una rama avanzada del aprendizaje automático, emplea redes neuronales multicapa (McCarthy *et al.*, 2020) y (McCarthy *et al.*, 2021). Estos modelos son especialmente adecuados para tratar datos de series temporales, como los precios de las criptomonedas, modelando patrones intrincados e interdependencias en los datos. En esta sección, realizamos una revisión de la literatura existente sobre la predicción de los precios de las criptomonedas utilizando técnicas de aprendizaje automático y aprendizaje profundo.

En las etapas iniciales, el modelo más reconocido fue el modelo autorregresivo de media móvil, ARIMA (Abu Bakar y Rosbi, 2017). Posteriormente, (Kim, Jun y Lee, 2021) utilizaron el modelo GARCH (modelo autorregresivo de heteroscedasticidad condicional) para pronosticar los

datos del mercado de criptomonedas. Si bien estas técnicas pueden utilizarse eficazmente para predecir a corto plazo

Sin embargo, no son adecuados para problemas no lineales y presentan un rendimiento de predicción a largo plazo deficiente (Fang *et al.*, 2023) y (Safari, Khalfalla e Imani, sin fecha). Para resolver este problema, se introdujo el aprendizaje automático para analizar series temporales y se ha aplicado con éxito a la predicción del precio de las criptomonedas. La capacidad del aprendizaje automático para procesar grandes y complejos volúmenes de datos ha resuelto muchas de las limitaciones de los métodos tradicionales. Los métodos de aprendizaje automático incluyen Support Vector Machine (SVM), árbol de decisión, ingenuo Bayes, bosque aleatorio (Rathan, Sai y Manikanta, 2019). (Orte *et al.*, 2023) combinaron árboles de decisión y modelos SVM para predecir las tendencias futuras de los precios. (Cortez, Rodríguez-García y Mongrut, 2020) desarrollaron una SVM ponderada por características y un algoritmo K-nearest neighbor para predecir el índice del mercado de criptodivisas. Los resultados experimentales han demostrado que el modelo tiene una buena capacidad de predicción a corto, medio y largo plazo (Rabiee y Safari, 2023) y (Safari, Khalfalla e Imani, 2022).

En los últimos años, se han utilizado técnicas de aprendizaje profundo que dependen únicamente de conjuntos de datos para predecir los precios de las criptomonedas, eliminando la necesidad de conocimientos expertos (Safari, Khalfalla e Imani, 2022). En consecuencia, su uso en la predicción de criptomonedas se ha convertido cada vez más en un punto central de la investigación académica. Los métodos de aprendizaje profundo abarcan la Unidad Recurrente Controlada (GRU), la RNN, la Red Neuronal Convolucional (CNN), la LSTM y la BiLSTM (Wegayehu y Muluneh, 2022), (Omran *et al.*, 2021), (Seabe, Moutsinga y Pindza, 2023). En, CNN se empleó secuencialmente para la predicción del precio de la criptodivisa (Ramadhani *et al.*, 2018; Alonso-Monsalve *et al.*, 2020). En (Seabe, Moutsinga y Pindza, 2023), se propuso un modelo Conv1D-LSTM, que fusiona CNN unidimensional y LSTM. Esta combinación aprovecha los puntos fuertes de ambas redes. (Ramakrishnan *et al.*, 2022) llevaron a cabo una investigación predictiva sobre la criptomoneda mundial

utilizando BiLSTM, lo que demuestra que BiLSTM posee una excelente precisión predictiva y una sólida capacidad de generalización.

El mecanismo de atención es un componente clave de la arquitectura de Transformer, ya que le proporciona la capacidad de centrarse en distintas partes de la secuencia de entrada a la hora de generar predicciones. Esta característica es especialmente beneficiosa en la predicción de series temporales, ya que permite al modelo ponderar de forma diferente la importancia de los eventos pasados y recientes, mejorando así significativamente la precisión de las predicciones. Los autores de (Totaro, Hussain y Scardapane, 2020) demuestran la aplicación de esta técnica en la predicción de series temporales con la red neuronal recurrente de doble etapa basada en la atención. Este modelo se aplica a los datos horarios del precio de Dogecoin para su predicción en el tiempo. El artículo (Zhang *et al.*, 2021) introduce un enfoque novedoso para predecir los precios de las criptomonedas. Los autores proponen un modelo que combina CNNs con canales de memoria ponderada y atenta. Esta combinación única permite al modelo extraer eficazmente características de los datos y capturar patrones complejos.

Los investigadores han propuesto adaptaciones como el Transformador de fusión temporal y el Transformador de series temporales para optimizar su arquitectura para datos de series temporales. (Lim *et al.*, 2021) presenta otra adaptación de Transformer, denominada Transformer de fusión temporal, que combina entradas de alta dimensión y diversas procedentes de múltiples fuentes para producir previsiones precisas e interpretables para varios horizontes temporales. El documento (Sridhar y Sanagavarapu, 2021) presenta un enfoque novedoso para predecir los precios de Dogecoin utilizando un Transformador de autoatención multicabezal. (Li et al., 2019) propone una variante novedosa de Transformer, llamada Transformer de series temporales, que mejora el rendimiento y la eficiencia de Transformer en tareas de predicción de series temporales. El documento (Son et al., 2022) presenta un enfoque novedoso para predecir los precios de las criptomonedas mediante el análisis de las tendencias de los medios sociales. El

artículo (You *et al.*, 2022) presenta un nuevo Transformer espaciotemporal diseñado para

predecir datos de series temporales a corto plazo de alta dimensión. Este modelo de enfoque aprovecha una ecuación de transformación de información espaciotemporal y un mecanismo de atención continua para mejorar la precisión de la predicción. El artículo (Tanwar y Kumar, 2022) explora la aplicación de técnicas avanzadas de aprendizaje profundo para predecir los precios de las criptodivisas. Específicamente examina el uso de transformadores junto con redes LSTM para analizar datos de series temporales financieras para criptodivisas como Ethereum y Bitcoin. El documento (Yunsi, Lahcen y Azzouz Mohamed, 2023) explora la aplicación de redes neuronales Transformer para predecir los precios de las criptodivisas. El documento (Du, Côté y Liu, 2023) propone un método novedoso basado en la mecanismo de autoatención que contiene dos bloques de autoatención enmascarados diagonalmente que aprenden los valores perdidos a partir de una combinación ponderada de dependencias temporales y de características.

Al examinar la bibliografía, es evidente que la estructura de red neuronal Transformer ha experimentado un aumento de uso recientemente, en comparación con los modelos de aprendizaje profundo anteriores. Esta arquitectura se diseñó para resolver tareas de secuencia a secuencia a la vez que maneja dependencias de largo alcance con facilidad. La principal ventaja de Transformer frente a arquitecturas anteriores, como las RNN, es su capacidad para gestionar dependencias de largo alcance. En términos sencillos, tiene una memoria más potente a la hora de recordar conexiones antiguas.

3) Metodología

Esta sección proporciona un examen en profundidad de la Red Neuronal Performer propuesta y sus componentes, que incluyen el Análisis Técnico, el Mecanismo de Atención Multicabezal en Transformer, Performer y BiLSTM. Presenta los indicadores técnicos como características estadísticas que ayudan en el análisis de datos financieros. El Mecanismo de Atención Multicabezal, elemento clave de la arquitectura de Transformer, se explica en términos de su

funcionalidad y fórmula subyacente. El Performer, una variante eficiente del transformador, se describe junto con su relevancia para el

modelo. Por último, la sección cubre LSTM y BiLSTM, que son tipos de RNN utilizados en la red feedforward del método propuesto, destacando su papel en la captura de dependencias temporales en los datos.

3.1) Análisis técnico

Los indicadores técnicos proporcionan una medida cuantificada de las condiciones y tendencias del mercado, que pueden servir como valiosas características de entrada para un modelo de aprendizaje profundo. Los indicadores técnicos, cuando se utilizan como características en modelos de aprendizaje profundo para la predicción del precio de Bitcoin, pueden mejorar significativamente las capacidades predictivas del modelo. Estos indicadores encapsulan tendencias y comportamientos clave del mercado, proporcionando un conjunto de datos rico y cuantificado del que el modelo puede aprender. Mediante la integración de indicadores, los modelos de aprendizaje profundo pueden captar relaciones complejas y no lineales en los datos que el análisis tradicional podría pasar por alto. Por ejemplo, un modelo de aprendizaje profundo puede utilizar estos indicadores para identificar patrones que preceden a los movimientos del mercado, permitiendo predicciones más precisas de los cambios de precios. El modelo puede aprender de datos históricos cómo ciertos valores de indicadores se correlacionan con tendencias al alza o a la baja en los precios de Bitcoin. Este aprendizaje permite al modelo anticipar movimientos similares cuando estos patrones de indicadores se repiten. Además, los indicadores técnicos pueden servir como herramienta de normalización, ayudando a escalar y transformar los datos de entrada en un formato más digerible para el modelo, mejorando así el proceso de entrenamiento. Al proporcionar una entrada normalizada, el modelo puede centrarse en los patrones subyacentes en lugar de confundirse por el ruido o las diferencias de escala en los datos de precios brutos (Wang *et al.*, 2023).

En este estudio se utilizan varios indicadores técnicos, como la media móvil simple, la media móvil exponencial, las bandas de Bollinger, el RSI y el índice del canal de materias primas

(CCI). También se proporcionan las fórmulas matemáticas correspondientes a estos indicadores.

1. **Media móvil simple (SMA):** La SMA se calcula sumando el precio de un instrumento a lo largo de varios periodos de tiempo y dividiendo después la suma por el número de periodos. La fórmula es la siguiente

$$SMA = \frac{1}{n} \sum_{i=1}^n P_i \quad (1)$$

donde P_i es el precio en el periodo i y n es el número de periodos.

2. **Media móvil exponencial (EMA):** La EMA es un tipo de media móvil ponderada que da más importancia a los datos más recientes. La fórmula es:

$$EMA_i = (P_i - EMA_{i-1}) \times k + EMA_{i-1} \quad (2)$$

donde P_i es el precio actual, EMA_{i-1} es el valor EMA del precio anterior, y k es

$$\frac{2}{(number\ of\ periods)}.$$

3. **Bandas de Bollinger (BB):** Las BB consisten en una banda media con dos bandas externas. La banda central es una media móvil simple, y las bandas exteriores son desviaciones estándar de la banda central. Las fórmulas son:

$$Middle\ Band = SMA(n) \quad (4)$$

$$Upper\ Band = SMA(n) + k \times std(n) \quad (5)$$

$$Lower\ Band = SMA(n) - k \times std(n) \quad (6)$$

donde n es el número de periodos, y k es una constante (normalmente 2).

4. **Índice de Fuerza Relativa (RSI):** El RSI compara la magnitud de las ganancias recientes con las pérdidas recientes en un intento de determinar las condiciones de sobrecompra y sobreventa de un instrumento. La fórmula es:

$$RSI = 100 - \frac{100}{1 + RS} \quad (7)$$

donde RS es la ganancia media en n periodos dividida por la pérdida media en k periodos.

5. **Índice del Canal de Materias Primas (ICC):** El CCI mide la diferencia entre la variación del precio de un valor y su variación media. Las lecturas positivas altas indican que los precios están muy por encima de su media, lo que es una muestra de fortaleza. La fórmula es:

$$CCI = \frac{(TP - MA)}{(0,015 \times D)} \quad (8)$$

donde TP es el precio típico $TP = \frac{(High+Low+Close)}{3}$, MA es la media móvil, y D es la desviación media.

3.2) Mecanismo de atención multicabezal

El núcleo del modelo Transformer es el mecanismo de atención multicabezal, fundamental para el tratamiento de datos secuenciales, especialmente en el procesamiento del lenguaje natural. Este mecanismo permite al modelo prestar atención simultáneamente a varios segmentos de la información de entrada y detectar patrones complejos e interdependencias. En este marco, la atención multicabezal desempeña un papel crucial a la hora de discernir conexiones complejas en distintos intervalos de una secuencia, lo que mejora la capacidad del modelo para identificar vínculos temporales y elementos significativos.

Los componentes del mecanismo de atención son:

- Consultas (Q): la información que se desea recuperar de la entrada.
- Claves (K): Proporcionan el contexto necesario para los elementos de la secuencia.
- Valores (V): Contienen los datos reales relacionados con cada elemento de la secuencia.

Durante la fase de atención, las consultas se cotejan con las claves para determinar la cantidad de datos que deben extraerse de cada valor. La puntuación de atención se calcula del siguiente modo.

$$Attention(Q, K, V) = Softmax \left(\frac{Q \cdot K^T}{\sqrt{d_k}} \right) \cdot V \quad (9)$$

En la fórmula dada, d_k denota el tamaño de los vectores clave. La raíz cuadrada de d_k , $\sqrt{d_k}$, se utiliza para escalar

los cálculos. Durante la autoatención, cuando las consultas, las claves y los valores proceden de secuencias idénticas, este escalado ayuda al modelo a distribuir de forma óptima la información entre los distintos elementos de la secuencia.

La atención multicabezal es un componente clave de la arquitectura Transformer, que permite al modelo centrarse simultáneamente en distintas partes de la secuencia de entrada. Cada cabeza de atención procesa la entrada de forma independiente, lo que permite al modelo captar diferentes tipos de información de la misma secuencia de entrada. Las siguientes fórmulas corresponden al mecanismo de Atención Multicabezal en los modelos Transformer. La función de Atención Multicabezal se define como sigue:

$$Multi - Head(Q, K, V) = Concat(head_1, \dots, head_h) W_o \quad (10)$$

En esta ecuación, cada $head_i$ representa una cabeza de atención individual, y W_o es la matriz de pesos de salida. La función *Concat* concatena la salida de todas las cabezas de atención, y este resultado se multiplica por la matriz de pesos de salida W_o .

Cada cabeza de atención individual ($head_i$) se calcula del siguiente modo:

$$)_{Ki}^T head_i = Attention(QW_{Qi}, KW_{Ki}, VW_{Vi}) = Softmax\left(\frac{(QW_{Qi}) \cdot (KW_{Ki})}{\sqrt{d_k}}\right) \times VW_{Vi} \quad (11)$$

En esta ecuación, QW_{Qi} , KW_{Ki} , y VW_{Vi} son las matrices de consulta, clave y valor para la i -ésima cabeza de atención, respectivamente. Estas matrices se obtienen multiplicando las matrices de consulta, clave y valor de entrada (Q , K y V) por sus correspondientes matrices de peso (QW_{Qi} , KW_{Ki} y VW_{Vi}). El

La función Softmax se aplica al producto punto de las matrices de consulta y clave, escalado por el cuadrado

raíz de la dimensionalidad de las claves $\sqrt{d_k}$. A continuación, este resultado se multiplica elemento a elemento por la matriz de valores VW_{Vi} . Estas fórmulas permiten al modelo procesar la entrada con múltiples cabezas de atención, cada una captar distintos tipos de información a partir de la misma secuencia de entrada. El diagrama de bloques del mecanismo de atención multicabezal se muestra en la Fig. 1.

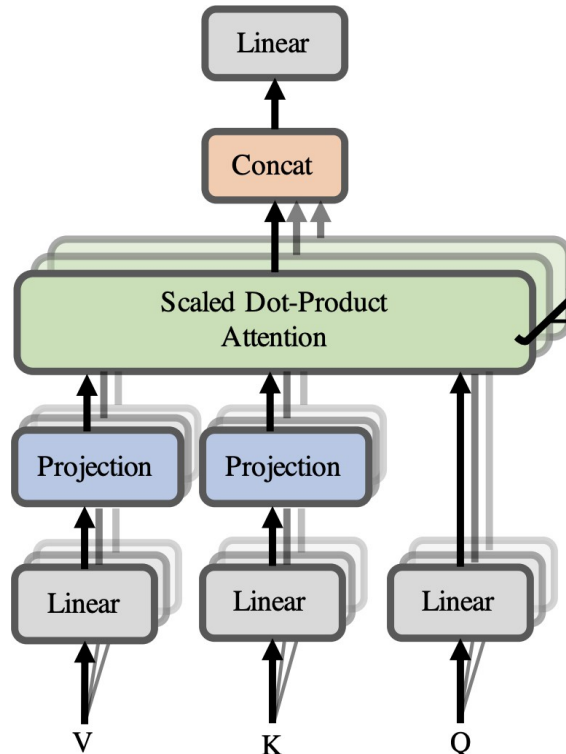


Fig. 1. Diagrama de bloques del mecanismo de atención multicabezal

3.3) Artista

El artículo "Rethinking Attention with Performers" presenta Performers, una nueva arquitectura de transformadores que puede estimar transformadores regulares de atención de rango completo con una precisión demostrable, pero con una complejidad de espacio y tiempo solo lineal (Wang *et al.*, 2023). Los Performers utilizan un nuevo enfoque llamado Fast Attention Via Positive Orthogonal Random features (FAVOR+), que también puede utilizarse para modelar eficientemente mecanismos de atención kernelizables más allá de Softmax. Esto permite comparar con precisión Softmax con otros kernels en tareas a gran escala, fuera del alcance de los Transformers normales. Los Performers se probaron en diversas tareas, desde la predicción de píxeles hasta modelos de texto y modelado de secuencias de proteínas, y demostraron resultados competitivos con otros métodos eficientes de atención dispersa y densa (Choromanski *et al.*, 2020).

Consideremos una serie temporal con longitud L . La atención estándar de producto punto, es un mapeo que

toma las matrices Q, K y $V \in \mathbb{R}^{L \times d}$ como entrada, donde d es la dimensión del estado oculto (la dimensión de la representación latente). Estas matrices Q, K y V son intermedias.

de las series temporales, y sus filas pueden verse como consultas, claves y valores de la estructura de datos del diccionario continuo, respectivamente. La atención bidireccional punto-producto adopta la siguiente forma, donde $A \in \mathbb{R}^{L \times L}$ es la denominada matriz de atención:

$$Attention_{\leftrightarrow}(Q, K, V) = D^{-1} AV, \quad (12)$$

$$A = \exp\left(\frac{Q \cdot K^T}{\sqrt{d_k}}\right), \quad (13)$$

$$D = \text{diag}(A \mathbf{1}_L), \quad (14)$$

aquí, la función $\exp(-)$ se aplica de forma elemental, $\mathbf{1}_L$ es un vector todo-uno de longitud L , y $\text{diag}(-)$

es una matriz diagonal con el vector de entrada como diagonal. La complejidad temporal y espacial de

que calculan (1) son $O(L^2 d)$ y $O(L^2 + Ld)$ respectivamente, porque A tiene que almacenarse explícitamente.

Por lo tanto, en principio, la atención punto-producto de tipo (1) es incompatible con el procesamiento de extremo a extremo de series temporales largas. La atención bidireccional se utiliza en la autoatención del codificador y la atención del codificador-decodificador en las arquitecturas Seq2Seq. Otro tipo significativo de atención es la atención punto-producto unidireccional, que adopta la forma:

$$Attention(Q, K, V) = D^{-1} AV, \quad (15)$$

$$A = tril(K) \quad (16)$$

$$D = diag(A^T_L), \quad (17)$$

donde $tril(-)$ devuelve la parte triangular inferior de la matriz del argumento, incluyendo la diagonal. La dirección

La atención unidireccional se utiliza para el modelado generativo autorregresivo, por ejemplo, como autoatención en los Transformadores generativos, así como en la parte decodificadora de los Transformadores Seq2Seq. FAVOR+ funciona

para bloques de atención utilizando matrices $A \in \mathbb{R}^{L \times L}$ de la forma $A(i, j) = K(q_i^T, k_j^T)$, con q/k_{ij}

que representa el i^{th} consulta/clave fila-vector en Q/K y núcleo $K: \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ definido para $+$

el mapeo (normalmente aleatorio): $\varphi: \mathbb{R}^d \rightarrow \mathbb{R}^d$ (para algún $r > 0$) como:

$$K(x, y) = \mathbb{E} [\varphi(x)^T \varphi(y)]. \quad (18)$$

Nos referimos a $\varphi(u)$ como un mapa de características aleatorio para $u \in \mathbb{R}^d$. Para $Q, K \in \mathbb{R}^L$ con filas dadas como $\varphi(q)_i^{TT}$ y $\varphi(k)_i^{TT}$ respectivamente, la ecuación (19) conduce directamente al mecanismo de atención eficiente de la forma:

$$Attention(Q, K, V) = D^{-1} V^T \quad (19)$$

$$\begin{aligned}
 &\leftrightarrow (Q \cdot (K) \cdot V), \\
 {}^{D^{\wedge}1} &= \mathit{diag} \left(Q \cdot \begin{matrix} ' & ' & T \\ & & \end{matrix} \cdot 1_L \right), \tag{20} \\
 &(K)
 \end{aligned}$$

La Fig. 2 muestra el diagrama de bloques del mecanismo FAVOR+.

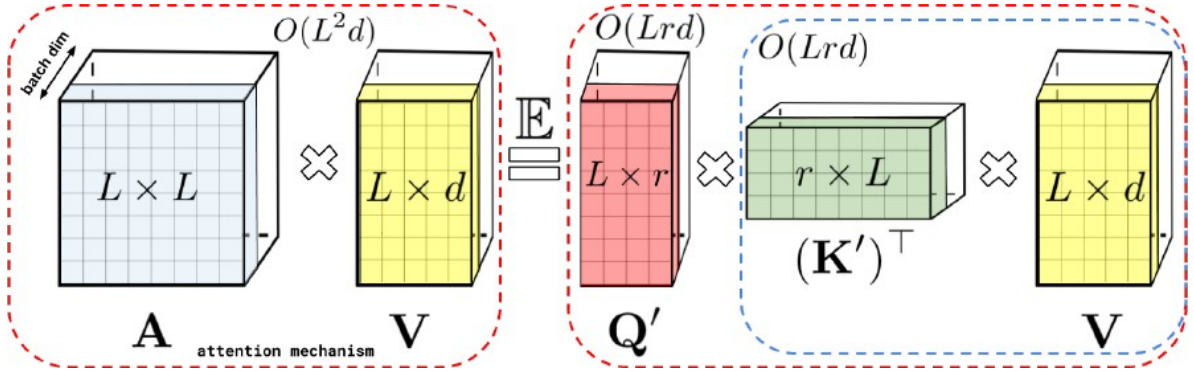


Fig. 2. Diagrama de bloques del mecanismo FAVOR+ (Choromanski *et al.*, 2020)

3.4 Memoria a corto plazo bidireccional (BiLSTM)

Las redes LSTM tienen una ventaja significativa sobre las RNN tradicionales debido a su capacidad para capturar dependencias a largo plazo en datos secuenciales. Las RNN tradicionales sufren el problema del gradiente evanescente, que les dificulta el aprendizaje y la propagación de la información contextual a través de secuencias largas. Las LSTM, que incorporan una célula de memoria y mecanismos de activación, abordan esta limitación con elegancia. Estas características permiten a las LSTM controlar y gestionar el flujo de información, decidiendo qué retener y qué descartar en diferentes escalas temporales. Esto hace que las LSTM sean especialmente eficaces en tareas en las que es crucial comprender el contexto y las relaciones entre las distintas partes de la secuencia de entrada, como en el procesamiento del lenguaje natural, el análisis de series temporales, etc. (Wang *et al.*, 2023). Por lo tanto, las LSTM suelen superar a las RNN tradicionales en tareas que implican secuencias largas y dependencias complejas. Las operaciones dentro de una unidad LSTM se rigen por las siguientes ecuaciones:

$$f = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f), \quad (21)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i), \quad (22)$$

$$\tilde{C}_t = \tanh(h(W_C \cdot [h_{t-1}, x_t] + b_C)), \quad (23)$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (24)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o), \quad (25)$$

$$h_t = o_t * \tanh(C_t), \quad (26)$$

Aquí f_t , i_t , o_t representan las activaciones de las puertas de olvido, entrada y salida, respectivamente; C_t es el estado de la celda; h_t es el estado oculto; x_t es la entrada en el paso de tiempo actual; σ denota la función de activación sigmoide; y W y b son los pesos y sesgos asociados a cada puerta. Estos permiten a las LSTM capturar eficazmente las dependencias temporales dentro de las secuencias de datos. El diagrama de bloques de LSTM se muestra en la Fig. 3.

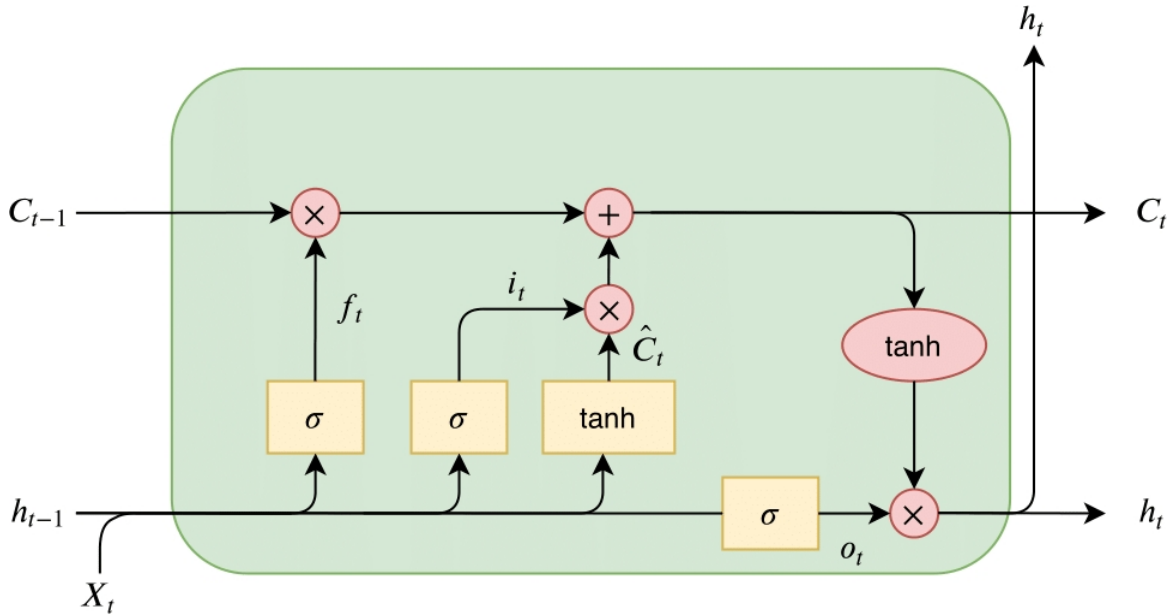


Fig. 3. Diagrama de bloques de LSTM

Las redes BiLSTM son una extensión de las redes LSTM tradicionales. Las BiLSTM procesan los datos en ambas direcciones con dos capas ocultas separadas que se envían a la misma capa de salida.

A diferencia de una LSTM tradicional, que procesa los datos secuencialmente desde el principio hasta el final de la secuencia, una BiLSTM procesa los datos en ambas direcciones. Una capa LSTM procesa la secuencia desde el principio hasta el final (hacia delante), mientras que la otra la procesa desde el final hasta el principio (hacia atrás). Este procesamiento bidireccional ayuda a captar patrones que podrían pasarse por alto con una LSTM unidireccional, ya que permite a la red captar información de estados futuros además de estados pasados. Matemáticamente, la salida de una BiLSTM en un determinado paso de tiempo t suele representarse como la concatenación del estado oculto anterior y el estado oculto posterior. Si denotamos el estado y estados ocultos hacia atrás en el tiempo como $t \rightarrow h_t$ y $\leftarrow h_t$ respectivamente, la salida y_t en el tiempo t puede ser computado como:

$$y = [h_t^{\rightarrow}, h_t^{\leftarrow}] \quad (27)$$

Esta salida y_t se pasa entonces a la siguiente capa de la red o se utiliza para calcular la predicción de el paso temporal actual. La capacidad de capturar tanto el contexto pasado (hacia atrás) como el futuro (hacia delante) convierte a BiLSTM en un potente modelo para tareas que requieren comprender todo el contexto de la secuencia de entrada, como el modelado del lenguaje, la generación de texto y la traducción automática (Wang *et al.*, 2023).

3.5) La red neuronal Performer propuesta

De acuerdo con las subsecciones anteriores, en esta subsección se presenta el Performer propuesto. En esta estructura, después de los bloques performer, se añaden dos capas BiLSTM y FC para aumentar el rendimiento de la predicción. El diagrama de bloques de LSTM se muestra en la Fig. 4.

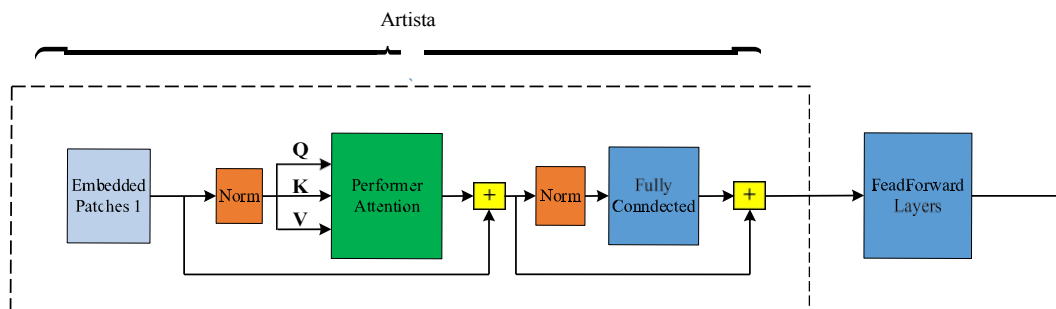


Fig. 4. Diagrama de bloques de la red neuronal Performer

En la estructura que ha mencionado, se añaden dos capas BiLSTM y FC (Fully Connected) después de los bloques Performer para aumentar el rendimiento de la predicción. La integración de BiLSTM en el feedforward del Performer mejora el rendimiento del modelo al combinar los puntos fuertes de ambos modelos. El Performer puede captar eficazmente las dependencias de largo alcance, mientras que el BiLSTM puede procesar eficazmente los datos secuenciales. Esta combinación permite al modelo manejar una gama más amplia de patrones de datos, mejorando su rendimiento predictivo. La ausencia de decodificador sugiere que este modelo está diseñado principalmente para tareas como la extracción de características, el aprendizaje de representaciones o la clasificación, más que para tareas de secuencia a secuencia que requieren un decodificador explícito.

4) Simulaciones

Esta sección profundiza en los intrincados procedimientos de entrenamiento y validación del modelo Performer propuesto. Este modelo está diseñado específicamente para predecir las fluctuaciones de los precios de Bitcoin, Ethereum y Litecoin desde enero de 2018 hasta marzo de 2024 en varios intervalos de tiempo, incluidos el diario y el horario. Los modelos utilizados para la comparación incluyen BiLSTM, Multi-head Attention Transformer con y sin indicadores técnicos, Performer, Performer integrado con BiLSTM. Estos modelos se evalúan utilizando cuatro métricas de rendimiento: Error cuadrático medio (MSE), R-cuadrado, Error cuadrático medio (RMSE) y Error logarítmico cuadrático medio (MSLE). Estas métricas se emplean para

demostrar el rendimiento comparativo de los modelos propuestos.

métodos. Además, las redes propuestas se yuxtaponen a otros métodos citados en el estado actual de la técnica, lo que proporciona una comparación exhaustiva de su eficacia.

4.1) Eficacia de los métodos de predicción de precios en términos de MSE

En esta subsección, se introduce *RMSE* para evaluar en profundidad el rendimiento de las redes neuronales. La ecuación (28) muestra la fórmula RMSE.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y})^2}, \quad (28)$$

donde y es el valor real y \hat{y} es el valor predicho, i es el índice de datos y n es la cantidad de datos. *RMSE* representa la raíz cuadrada del segundo momento muestral de las diferencias entre valores predichos y valores observados o la media cuadrática de estas diferencias. Mide la magnitud media de los errores sin tener en cuenta su dirección.

Los modelos evaluados para el marco temporal diario y horario de Bitcoin son LSTM (Awoke et al., 2021), Stochastic Neural Network (Jay et al., 2020), Hybrid LSTM and GRU (Girsang and Stanley, 2023), BiLSTM, Multi-head Transformer con y sin indicadores técnicos, Performer, y Performer combinado con BiLSTM. Las Figs. 5-6 muestran el rendimiento del modelo en términos de la métrica RMSE.

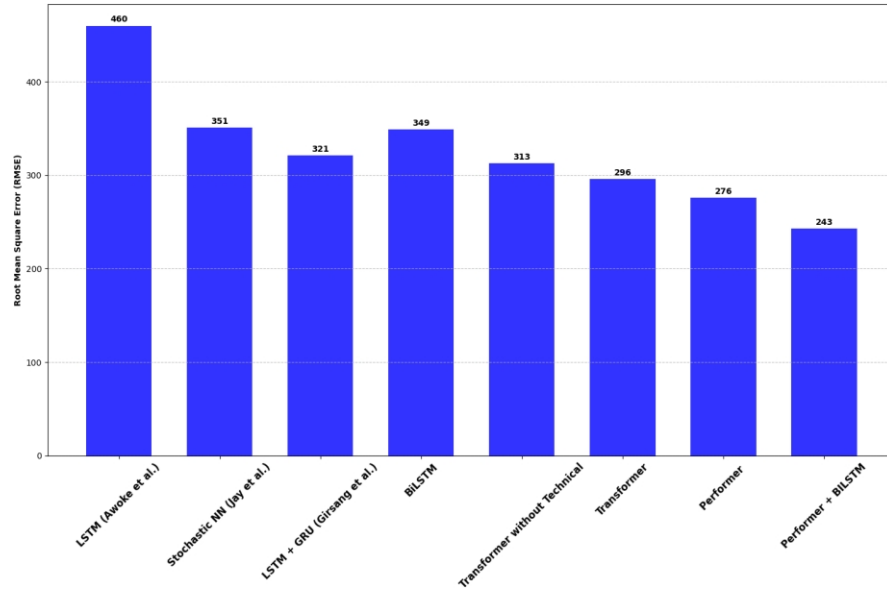


Fig. 5. Comparación de los valores de RMSE para la predicción horaria del precio del BTCUSD utilizando los métodos más avanzados y los métodos basados en transformadores propuestos.

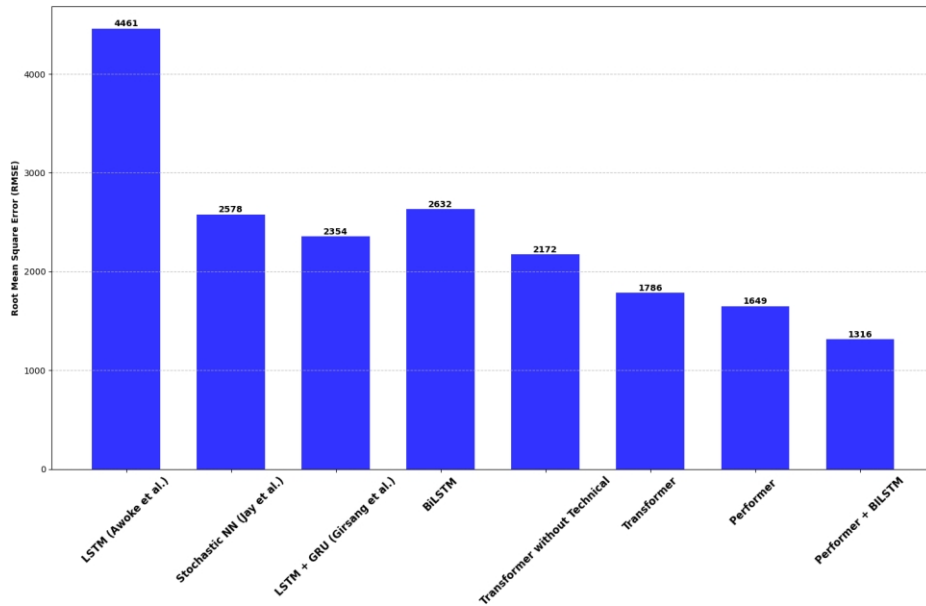


Fig. 6. Comparación de los valores de RMSE para la predicción del precio diario del BTCUSD utilizando los métodos más avanzados y los métodos basados en Transformer propuestos.

Basándose en las Figuras 5 y 6, puede deducirse que la red neuronal propuesta basada en Transformer supera a otros métodos de aprendizaje profundo. El Transformador de atención multicabezal es mejor que BiLSTM porque puede capturar dependencias en los datos independientemente de su posición en la secuencia. Esto es particularmente útil en tareas en las que la posición de los puntos de datos no es tan

importantes como su relación entre sí. Por otra parte, BiLSTM, al ser un modelo recurrente, procesa los datos secuencialmente, lo que puede hacerlo menos eficiente a la hora de captar dependencias de largo alcance. Además, los indicadores técnicos proporcionan información valiosa sobre las tendencias y pautas del mercado. Cuando se utilizan como extractores de características, pueden mejorar la capacidad del modelo para captar y aprender de estos patrones, lo que se traduce en un mejor rendimiento. Estos indicadores pueden proporcionar información sobre diversos aspectos del mercado, como la dirección de la tendencia, la volatilidad, el impulso y la fuerza del mercado, que son cruciales para predecir los movimientos de los precios. Al incorporar esta información al modelo, podemos mejorar su capacidad para hacer predicciones precisas.

En el modelo Transformer original, el mecanismo de atención multicabezal calcula una suma ponderada de todos los elementos de entrada para cada elemento de salida. Esto requiere una cantidad cuadrática de cálculo y memoria con respecto a la longitud de la secuencia, lo que dificulta el procesamiento de secuencias largas. El Performer, en cambio, utiliza una técnica llamada Random Feature Maps para aproximarse al mecanismo de atención. Esta técnica permite al Performer calcular el mecanismo de atención en tiempo lineal, lo que reduce significativamente el coste computacional y permite al modelo escalar a secuencias mucho mayores. Además, el Performer mantiene un nivel de expresividad similar al del Transformer original. Puede modelar patrones complejos en los datos y capturar dependencias de largo alcance entre elementos de la secuencia, lo que resulta crucial para muchas tareas como el modelado del lenguaje y la predicción de series temporales.

La integración de BiLSTM en el feedforward del Performer mejora el rendimiento del modelo al combinar los puntos fuertes de ambos modelos. El Performer puede captar eficazmente las dependencias de largo alcance, mientras que el BiLSTM puede procesar eficazmente los datos secuenciales. Esta combinación permite al modelo manejar una gama más

amplia de patrones de datos, mejorando su rendimiento predictivo. En conclusión, la figura sugiere que el modelo Performer + BiLSTM es el mejor

para la predicción del precio diario y horario del Bitcoin, ya que tiene el RMSE más bajo y puede captar tanto la atención como las dependencias bidireccionales de los datos.

4.2) Evaluación del rendimiento de los modelos basados en transformadores en la predicción del precio de las criptomonedas

En esta subsección, nos embarcamos en una exploración exhaustiva en términos de MSE, R^2 , RMSE y

MSLE para la predicción del precio de cierre de Bitcoin, Ethereum y Litecoin en datos de marco temporal diario y horario. Equ. (29) representa la fórmula MSE.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2, \quad (29)$$

El MSE se calcula como la suma de las diferencias al cuadrado entre los valores reales y los predichos, dividida por el número de puntos de datos. En concreto, es la media de estas discrepancias al cuadrado en todas las observaciones. El MSE sirve como criterio de optimización de las redes neuronales durante la fase de entrenamiento, guiando el ajuste de los parámetros del modelo para minimizar los errores de predicción. El

² R fórmula viene dada por (30).

$$^2 R = 1 - \frac{Var(y - \hat{y})}{Var(y)}, \quad (30)$$

donde R^2 Mide la proporción de la varianza de la variable dependiente que es predecible a partir de las variables independientes. Proporciona una indicación de la bondad del ajuste y, por tanto, una medida

de la probabilidad de que las muestras no vistas sean predichas por el modelo. El *MSLE* se define como sigue.

$$MSLE = \frac{1}{n} \sum_{i=1}^n (\log(y_i + 1) - \log(\hat{y}_i + 1))^2, \quad (31)$$

MSLE Mide la relación entre los valores verdaderos y los predichos. Se aplica una transformación logarítmica para que los errores en la predicción de valores grandes y pequeños se traten proporcionalmente.

Las figuras (7-8) muestran la eficacia de Transformer Multi-head, Performer y Performer combinado con BiLSTM en la predicción de los precios de cierre diarios y horarios de Bitcoin. Las Tablas 1-2 presentan una evaluación exhaustiva del rendimiento de todos los modelos analizados en la subsección anterior.

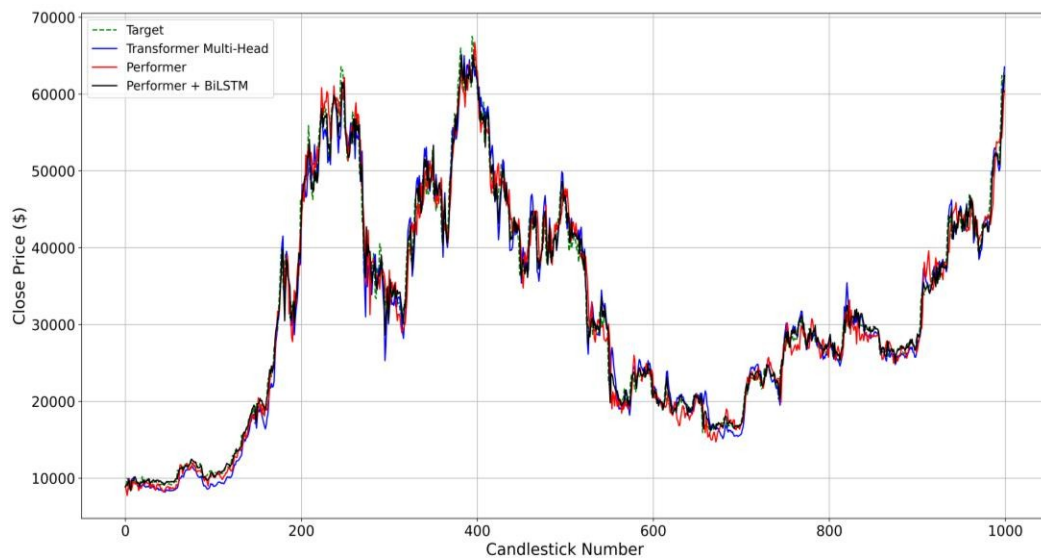


Fig. 7. Predicción del precio diario del BTCUSD utilizando Transformer Multi-head, Performer, y Performer + BiLSTM.

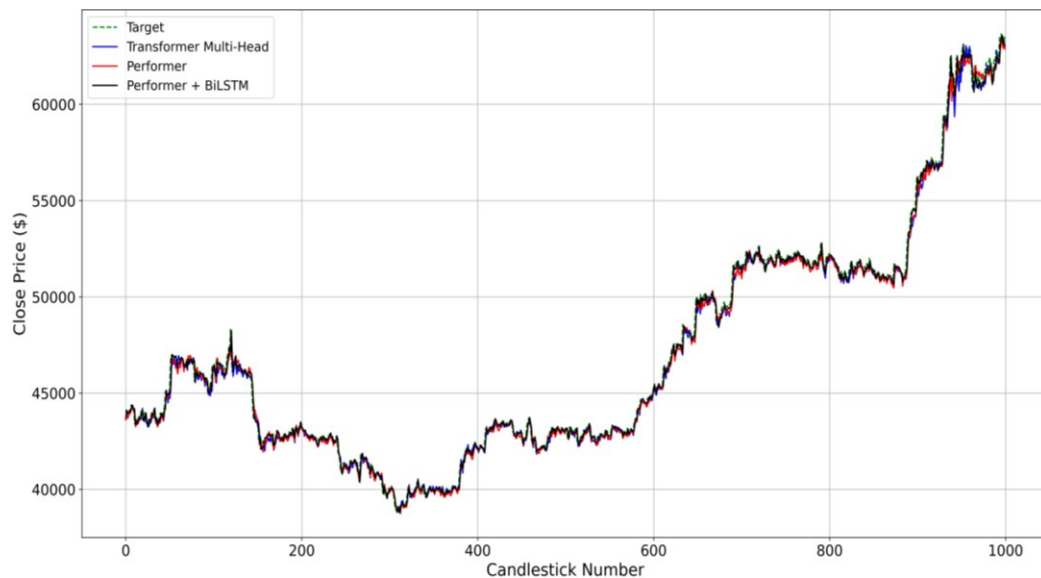


Fig. 8. Predicción del precio horario del BTCUSD utilizando Transformer Multi-head, Performer, y Performer + BiLSTM.

Tabla 1. Índices de rendimiento para la predicción del precio del BTCUSD horario utilizando algunos métodos del estado de la técnica y los métodos propuestos basados en transformadores

Método	MSE	RMSE	R-Cuadrado	MSLE
LSTM (Awoke <i>et al.</i> , 2021)	195426	442	0.9971	0.00124
NN estocástica (Jay <i>et al.</i> , 2020)	121802	351	0.9991	0.00076
LSTM+GRU (Girsang y Stanley, 2023)	103041	321	0.9992	0.00069
BiLSTM	122312	349	0.9991	0.00075
Transformador multicabezal sin técnica Indicador	98433	313	0.9994	0.00051
Transformador multicabezal	87710	296	0.9996	0.00032
Artista	77105	276	0.9997	0.00022
Intérprete + BiLSTM	59481	243	0.9998	0.00012

Tabla 2. Índices de rendimiento para la predicción del precio diario del BTCUSD utilizando algunos métodos del estado de la técnica y los métodos propuestos basados en transformadores

Método	MSE	RMSE	R-Cuadrado	MSLE
LSTM (Awoke <i>et al.</i> , 2021)	19900521	4461	0.897	0.0186
NN estocástica (Jay <i>et al.</i> , 2020)	6646084	2578	0.918	0.0179
LSTM+GRU (Girsang y Stanley, 2023)	5541316	2354	0.921	0.0172
BiLSTM	6929056	2632	0.917	0.0179
Transformador multicabezal sin técnica Indicador	4719246	2172	0.932	0.0167
Transformador multicabezal	3189856	1786	0.988	0.014
Artista	2720880	1649	0.99	0.008
Intérprete + BiLSTM	1732590	1316	0.993	0.004

Los resultados de las Tablas (1-2) y las Figs. (7-8) muestran que el modelo Performer + BiLSTM supera a los otros modelos en la predicción del precio de cierre de Bitcoin tanto para datos diarios como horarios. El rendimiento del método propuesto para predecir el precio de Ethereum se muestra en las Figs. (9-10) y las Tablas (3-4).

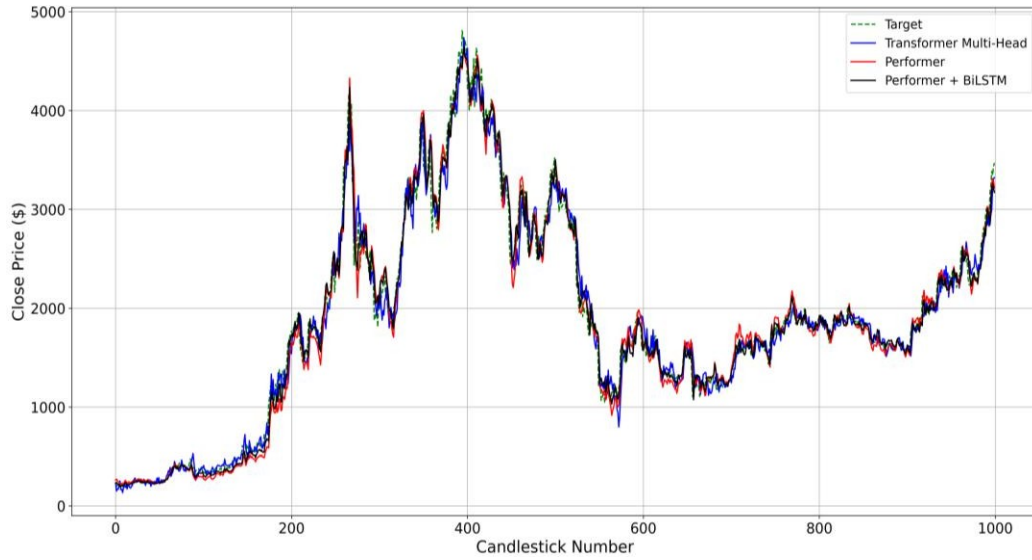


Fig. 9. Predicción del precio diario del ETHUSD utilizando Transformer Multi-head, Performer, y Performer + BiLSTM.

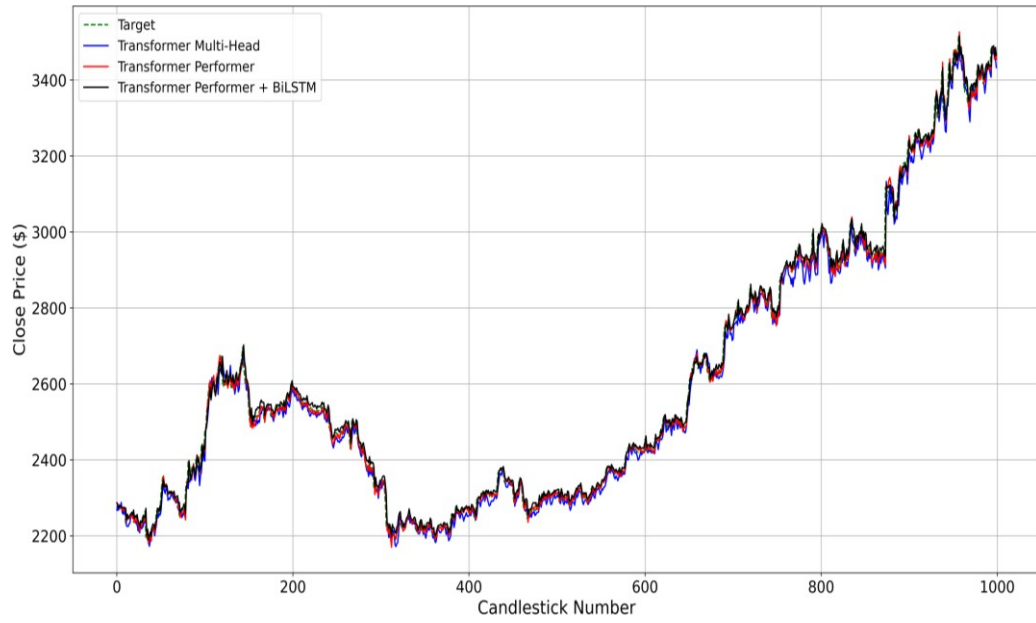


Fig. 10. Predicción del precio de ETHUSD por hora utilizando Transformer Multi-head, Performer, y Performer + BiLSTM.

Tabla 3 Índices de rendimiento para la predicción del precio del ETHUSD horario utilizando algunos métodos del estado de la técnica y los métodos propuestos basados en transformadores

Método	MSE	RMSE	R-Cuadrado	MSLE
LSTM (Awoke <i>et al.</i> , 2021)	934	30	0.9987	0.00052
NN estocástica (Jay <i>et al.</i> , 2020)	798	28.2	0.9990	0.00049
LSTM+GRU (Girsang y Stanley, 2023)	765	27.6	0.9991	0.00047
BiLSTM	789	28	0.9990	0.00048
Transformador multicabezal sin técnica Indicador	668	26	0.9993	0.00042
Transformador multicabezal	518	22.7	0.9995	0.00036
Artista	477	21.8	0.9996	0.00022
Intérprete + BiLSTM	386	18.3	0.9997	0.00016

Tabla 4. Índices de rendimiento para la predicción del precio del ETHUSD diario utilizando algunos métodos del estado de la técnica y los métodos propuestos basados en transformadores

Método	MSE	RMSE	R-Cuadrado	MSLE
LSTM (Awoke <i>et al.</i> , 2021)	38342	195	0.9698	0.048
NN estocástica (Jay <i>et al.</i> , 2020)	30153	173	0.9780	0.040
LSTM+GRU (Girsang y Stanley, 2023)	28850	170	0.9788	0.037
BiLSTM	29601	172	0.9782	0.039
Transformador multicabezal sin técnica Indicador	24807	157	0.9820	0.035
Transformador multicabezal	14957	122	0.9890	0.023
Artista	10991	105	0.9919	0.011
Intérprete + BiLSTM	10017	100	0.9926	0.010

Los resultados de las Tablas (3-4) y las Figs. (9-10) muestran que el modelo Performer + BiLSTM supera a los otros modelos en la predicción del precio de cierre de Ethereum tanto para datos diarios como horarios. El rendimiento del método propuesto para predecir el precio de Ethereum se muestra en las Figs. (11-12) y las Tablas (5-6).

El modelo Performer + BiLSTM obtiene mejores resultados que los demás modelos en la predicción del precio de cierre de Ethereum, tanto para el marco temporal diario como para el horario. Las Figs. (11-12) y las Tablas (5- 6) presentan el rendimiento del método propuesto para predecir el precio de Litecoin.

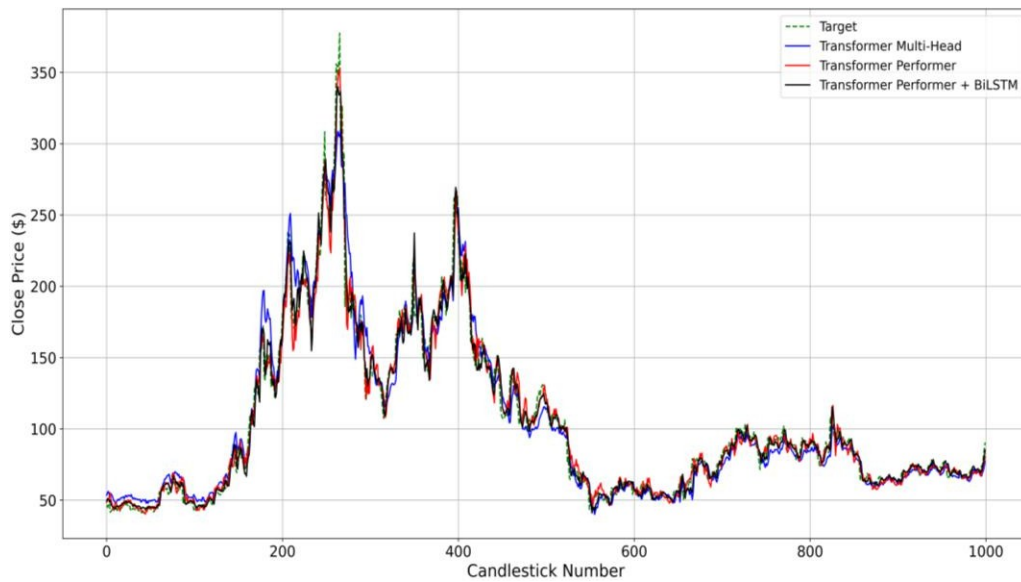


Fig. 11. Predicción del precio diario del LTCUSD utilizando Transformer Multi-head, Performer, y Performer + BiLSTM

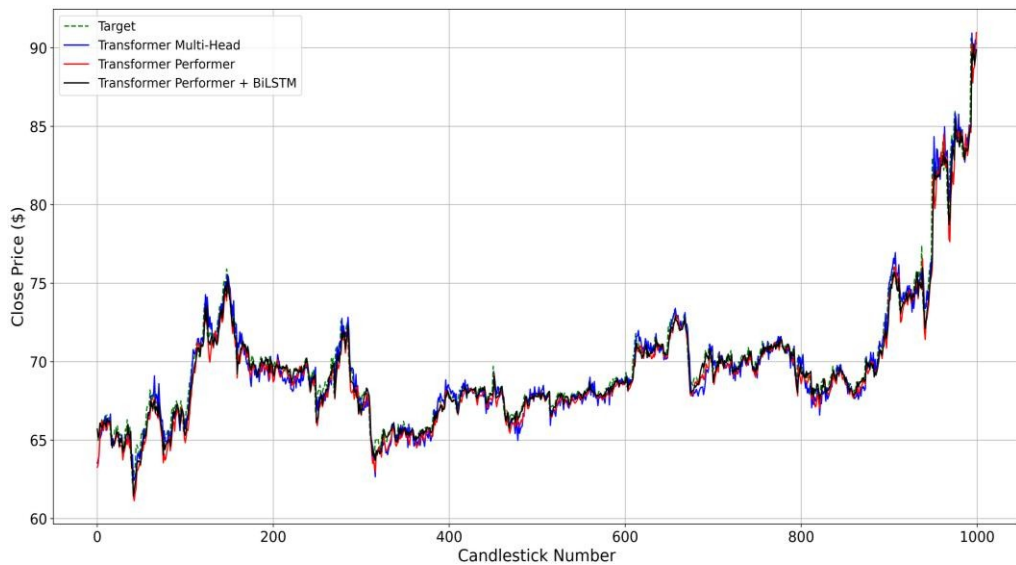


Fig. 12. Predicción del precio horario del LTCUSD utilizando Transformer Multi-head, Performer, y Performer + BiLSTM

Tabla 5. Índices de rendimiento para la predicción del precio del LTCUSD horario utilizando algunos métodos del estado de la técnica y los métodos propuestos basados en transformadores

Método	MSE	RMSE	R-Cuadrado	MSLE
LSTM (Awoke <i>et al.</i> , 2021)	9.45	3.07	0.9960	0.0067
NN estocástica (Jay <i>et al.</i> , 2020)	7.74	2.78	0.9973	0.0053
LSTM+GRU (Girsang y Stanley, 2023)	7.20	2.68	0.9976	0.00053
BiLSTM	7.85	2.80	0.9972	0.00057
Transformador multicabezal sin técnica	5.79	2.40	0.9980	0.00045
Indicador				
Transformador multicabezal	4.96	2.22	0.9984	0.00038
Artista	4.17	2.04	0.9987	0.00021
Intérprete + BiLSTM	3.70	1.92	0.9988	0.00020

Tabla 6. Índices de rendimiento para la predicción del precio del LTCUSD diario utilizando algunos métodos del estado de la técnica y los métodos propuestos basados en transformadores

Método	MSE	RMSE	R-Cuadrado	MSLE
LSTM (Awoke <i>et al.</i> , 2021)	195	14	0.909	0.0192
NN estocástica (Jay <i>et al.</i> , 2020)	185	13.6	0.918	0.0179
LSTM+GRU (Girsang y Stanley, 2023)	180	13.4	0.924	0.0177
BiLSTM	186	13.6	0.917	0.0179
Transformador multicabezal sin técnica	167	12.9	0.932	0.0167
Indicador				
Transformador multicabezal	134	11.5	0.954	0.0128
Artista	65	8.1	0.980	0.0061
Intérprete + BiLSTM	57	7.5	0.981	0.0057

Los datos de las Tablas 5-6 y las Figuras 11-12 demuestran que el modelo Performer combinado con BiLSTM supera a otros modelos en la previsión del precio de cierre de Litecoin tanto en el marco temporal diario como en el horario.

5) Conclusión

Esta investigación ha introducido una metodología novedosa para predecir series temporales de criptomonedas, concretamente Bitcoin, Ethereum y Litecoin. El enfoque combina el uso de indicadores técnicos, una red neuronal Performer y BiLSTM para captar la dinámica temporal y extraer características significativas de los datos brutos de criptomonedas. La aplicación de indicadores técnicos, como el RSI y el SMA, ha permitido extraer patrones y tendencias complejos que de otro modo podrían pasarse por alto. La red neuronal Performer, que utiliza el FAVOR+, ha demostrado ser más eficiente desde el punto de vista computacional y escalable que el mecanismo tradicional de atención multicabezal utilizado en los modelos Transformer. Además, la incorporación de BiLSTM en la red feedforward ha mejorado la capacidad del modelo para captar la dinámica temporal de los datos, procesándolos tanto hacia delante como hacia atrás. Esto es especialmente beneficioso para los datos de series temporales, en los que los puntos de datos pasados y futuros pueden influir en el estado actual. Las capas totalmente conectadas han permitido al modelo aprender complejas relaciones no lineales entre las características, equipándolo para manejar la complejidad y volatilidad que a menudo se observa en los movimientos de los precios de las criptomonedas. El método propuesto se ha aplicado a los marcos temporales diario y horario de las principales criptomonedas, como Bitcoin, Ethereum y Litecoin, y su rendimiento se ha comparado con el de otros métodos documentados en la literatura. Los resultados han demostrado el potencial del método propuesto para superar el rendimiento predictivo de los modelos existentes, lo que supone un avance significativo en el campo de la predicción del precio de las criptomonedas.

En futuros trabajos, pretendemos perfeccionar aún más el modelo y ampliar su aplicabilidad a otros mercados financieros y a diversos tipos de criptomonedas. La evolución continua de esta investigación podría conducir a modelos predictivos más precisos y eficientes en el análisis financiero. También tenemos previsto explorar la integración de diversas características, como el análisis de opiniones, la actividad en las redes sociales y las noticias.

titulares. Además, pretendemos mejorar nuestro método optimizando la arquitectura de la red, afinando los hiperparámetros y aplicando técnicas de regularización para mitigar el sobreajuste.

6) Referencia

Abu Bakar, N. y Rosbi, S. (2017) "Autoregressive integrated moving average (ARIMA) model for forecasting cryptocurrency exchange rate in high volatility environment: A new insight of bitcoin transaction", *International Journal of Advanced Engineering Research and Science*, 4(11), pp. 130-137.

Aghashahi, M. y Bamdad, S. (2023) "Analysis of different artificial neural networks for Bitcoin price prediction", *International Journal of Management Science and Engineering Management*. Taylor & Francis, 18(2), pp. 126-133. doi: 10.1080/17509653.2022.2032442.

Alonso-Monsalve, S. *et al.* (2020) 'Convolution on neural networks for high-frequency trend prediction of cryptocurrency exchange rates using technical indicators', *Expert Systems with Applications*. Elsevier, 149, p. 113250.

Awoke, T. *et al.* (2021) *Bitcoin Price Prediction and Analysis Using Deep Learning Models, Lecture Notes in Networks and Systems*. Springer Singapore. doi: 10.1007/978-981-15-5397-4_63.

Choromanski, K. *et al.* (2020) 'Rethinking attention with performers', *arXiv preprint arXiv:2009.14794*.

Cortez, K., Rodríguez-García, M. del P. y Mongrut, S. (2020) 'Exchange market liquidity prediction with the K-nearest neighbor approach: Crypto vs. fiat currencies', *Mathematics*. MDPI, 9(1), p. 56.

Du, W., Côté, D. y Liu, Y. (2023) 'Saits: Self-attention-based imputation for time series', *Sistemas expertos con aplicaciones*. Elsevier, 219, p. 119619.

Fang, Z. *et al.* (2023) 'Application of Non-Destructive Test Results to Estimate Rock

Mechanical Characteristics-A Case Study', *Minerals*. MDPI, 13(4), p. 472.

Girsang, A. S. y Stanley (2023) 'Hybrid LSTM and GRU for Cryptocurrency Price Forecasting Based on Social Network Sentiment Analysis Using FinBERT', *IEEE Access*. IEEE, 11(noviembre), pp. 120530-120540. doi: 10.1109/ACCESS.2023.3324535.

Goutte, S. *et al.* (2023) "Deep learning and technical analysis in cryptocurrency market", *Finance Research Letters*. Elsevier, 54, p. 103809.

Haryono, A. T., Sarno, R. y Sungkono, K. R. (2023) 'Transformer-Gated Recurrent Unit Method for Predicting Stock Price Based on News Sentiments and Technical Indicators', *IEEE Access*, 11(julio), pp. 77132-77146. doi: 10.1109/ACCESS.2023.3298445.

Jay, P. *et al.* (2020) "Stochastic neural networks for cryptocurrency price prediction", *IEEE Access*, 8, pp. 82804-82818. doi: 10.1109/ACCESS.2020.2990659.

Kim, J.-M., Jun, C. y Lee, J. (2021) "Forecasting the volatility of the cryptocurrency market by GARCH and Stochastic Volatility", *Mathematics*. MDPI, 9(14), p. 1614.

Labba Khaniki, M. A., Mirzaeibonehkhater, M. y Manthouri, M. (2023) 'Enhancing Pneumonia Detection using Vision Transformer with Dynamic Mapping Re-Attention Mechanism', en *2023 13th International Conference on Computer and Knowledge Engineering (ICCKE)*, pp. 144-149. doi: 10.1109/ICCKE60553.2023.10326313.

Li, S. *et al.* (2019) 'Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting', *Advances in neural information processing systems*, 32.

Lim, B. *et al.* (2021) "Temporal fusion transformers for interpretable multi-horizon time series forecasting", *International Journal of Forecasting*. Elsevier, 37(4), pp. 1748-1764.

McCarthy, P. X. *et al.* (2020) 'Long-term trends of diversity online', *arXiv e-prints*, p. arXiv-2003.

McCarthy, P. X. *et al.* (2021) 'Evolution of diversity and dominance of companies in online activity', *Plos one*. Public Library of Science, 16(4), p. e0249993.

Mohammadabadi, S. M. S. *et al.* (2023) 'Speed Up Federated Learning in Heterogeneous Environment: A Dynamic Tiering Approach', *arXiv preprint arXiv:2312.05642*.

- Omran, N. F. *et al.* (2021) 'Applying Deep Learning Methods on Time-Series Data for Forecasting COVID-19 in Egypt, Kuwait, and Saudi Arabia', *Complexity*, 2021. doi: 10.1155/2021/6686745.
- Orte, F. *et al.* (2023) "A random forest-based model for crypto asset forecasts in futures markets with out-of-sample prediction", *Research in International Business and Finance*. Elsevier, 64, p. 101829.
- Patwardhan, N., Marrone, S. y Sansone, C. (2023) 'Transformers in the Real World: A Survey on NLP Applications', *Information*. MDPI, 14(4), p. 242.
- Pichaiyuth, P. *et al.* (2023) "Price Trend Forecasting of Cryptocurrency Using Multiple Indicadores técnicos y SHAP", en *2023 20th International Joint Conference on Computer Science and Software Engineering (JCSSE)*. IEEE, pp. 150-154.
- Rabiee, P. y Safari, A. (2023) 'Safe Exploration in Reinforcement Learning: Training Backup Control Barrier Functions with Zero Training Time Safety Violations', *arXiv preprint arXiv:2312.07828*.
- Rahali, A. y Akhloufi, M. A. (2023) 'End-to-end transformer-based models in textual-based NLP', *AI*. MDPI, 4(1), pp. 54-110.
- Ramadhani, A. M. *et al.* (2018) 'Bitcoin price forecasting using neural decomposition and deep learning', *Journal of the Korea Industrial Information Systems Research*. Sociedad Coreana de Sistemas de Información Industrial, 23(4), pp. 81-92.
- Ramakrishnan, R. *et al.* (2022) 'Data-Driven Volatile Cryptocurrency Price Forecasting via Variational Mode Decomposition and BiLSTM', en *International Conference on Innovative Informática y Comunicaciones: Actas de ICICC 2022, Volumen 1*. Springer, pp. 651-663.
- Rathan, K., Sai, S. V. y Manikanta, T. S. (2019) 'Crypto-currency price prediction using decision tree and regression techniques', en *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)*. IEEE, pp. 190-194.
- Safari, K., Khalfalla, S. and Imani, F. (2022) 'Dependency Evaluation of Defect Formation and Printing Location in Additive Manufacturing', en *ASME International Mechanical Engineering Congress and Exposition*. American Society of Mechanical Engineers, p. V02AT02A016.

Safari, K., Khalfalla, S. e Imani, F. (sin fecha) 'Physics-Guided Deep Learning for Discovering and Monitoring Melt Pool Dynamics in Additive Manufacturing', *Disponibile en SSRN 4280300*.

Samii, A. *et al.* (2023) 'Comparison of DEEP-LSTM and MLP Models in Estimation of Evaporation Pan for Arid Regions', *Journal of Soft Computing in Civil Engineering*, 7(2).

Seabe, P. L., Moutsinga, C. R. B. y Pindza, E. (2023) "Forecasting cryptocurrency prices using LSTM, GRU, and bi-directional LSTM: a deep learning approach", *Fractal and Fractional*. MDPI, 7(2), p. 203.

Son, Y. *et al.* (2022) "Using Transformers and Deep Learning with Stance Detection to Forecast Cryptocurrency Price Movement", en *2022 13th International Conference on Information and Communication Technology Convergence (ICTC)*. IEEE, pp. 1-6.

Sridhar, S. y Sanagavarapu, S. (2021) 'Multi-head self-attention transformer for dogecoin price prediction', en *2021 14th International Conference on Human System Interaction (HSI)*. IEEE, pp. 1-6.

Tanwar, A. y Kumar, V. (2022) 'Prediction of Cryptocurrency prices using Transformers and Long Short term Neural Networks', en *2022 International Conference on Intelligent Controller and Computing for Smart Power (ICICCSP)*. IEEE, pp. 1-4.

Totaro, S., Hussain, A. y Scardapane, S. (2020) "A non-parametric softmax for improving neural attention in time-series forecasting", *Neurocomputing*. Elsevier, 381, pp. 177-185.

Vaswani, A. *et al.* (2017) 'Attention is all you need', *Advances in neural information processing systems*, 30. (en inglés).

Wang, Z. *et al.* (2023) 'Learning-based stock trending prediction by incorporating technical indicators and social media sentiment', *Cognitive Computation*. Springer, 15(3), pp. 1092-1102.

Wegayehu, E. B. y Muluneh, F. B. (2022) 'Short-Term Daily Univariate Streamflow Forecasting Using Deep Learning Models', *Advances in Meteorology*, 2022. doi: 10.1155/2022/1860460.

You, Y. *et al.* (2022) "Spatiotemporal Transformer Neural Network for Time-Series Forecasting", *Entropy*. MDPI, 24(11), p. 1651.

Yunsi, G., Lahcen, L. y Azzouz Mohamed, D. A. A. (2023) 'CRYPTOCURRENCY PRICE FORECASTING USING TRANSFORMER MODEL'.

Zhang, Z. *et al.* (2021) 'Forecasting cryptocurrency price using convolutional neural networks with weighted and attentive memory channels', *Expert Systems with Applications*. Elsevier, 183, p. 115378.

Zhao, H., Crane, M. y Bezbradica, M. (2022) "¡Atención! Transformador con Sentimiento en Cryptocurrencies Price Prediction', *Conferencia Internacional sobre Complejidad, Sistemas de Información del Futuro y Riesgo, COMPLEXIS - Actas*, 2022-abril, pp. 98-104. doi: 10.5220/0011103400003197.