



Facultad de Ciencias Económicas y Empresariales

Predicción de la cotización de acciones en bolsa mediante Deep Learning

Autor: Ainhoa Gallego Varela

Director: Jenny Alexandra Cifuentes Quintero

MADRID | Junio 2023

Resumen

La predicción precisa de los precios de las acciones a corto plazo desempeña un papel crucial en los mercados financieros, ya que brinda a los inversores la capacidad de tomar decisiones informadas y maximizar sus rendimientos. Aunque previamente se creía que las series de tiempo de los precios de las acciones eran impredecibles, investigaciones recientes han revelado la existencia de patrones y tendencias que permiten realizar predicciones con cierta precisión. En este contexto, las técnicas de aprendizaje profundo se han destacado como una poderosa herramienta, ya que son capaces de capturar patrones complejos y adaptarse a los cambios en los datos. Con el objetivo de aprovechar estas ventajas, este trabajo de fin de grado se enfoca en realizar una comparativa de estrategias basadas en aprendizaje profundo para predecir a corto plazo el precio de cierre de las acciones de Acciona, una empresa de gran relevancia en el índice bursátil IBEX 35. Mediante el análisis exhaustivo y la evaluación de diferentes técnicas y modelos, se busca mejorar la capacidad de predicción.

Para ello, se realiza la descarga y preprocesamiento de los datos históricos de los precios de cierre de las acciones de Acciona desde 2015 hasta 2023. Durante este proceso, se lleva a cabo la normalización de los datos y se verifica la ausencia de valores faltantes para garantizar la integridad de la serie temporal. Posteriormente, se realiza la división de la serie temporal en conjuntos de entrenamiento y prueba, asignando un 60 % de los datos para el conjunto de entrenamiento y un 40 % para el conjunto de pruebas. Además, se implementa la estrategia de “early stopping” con el objetivo de mejorar el rendimiento del modelo y evitar el sobreajuste. Dentro de la implementación, se explora la utilización de diferentes tamaños de ventana, incluyendo opciones de 30, 60 y 90 días, con diferentes alternativas de tamaño de muestra, como 16, 32 y 64. Estas variaciones en los tamaños de ventana y muestra permiten evaluar la influencia de estos parámetros en la capacidad de predicción del modelo y determinar la combinación óptima para obtener resultados precisos. Tras implementar y comparar diferentes técnicas de arquitecturas de redes neuronales, se concluye que el modelo más ajustado a los datos históricos de los precios de las acciones de Acciona es aquel basado en una arquitectura CNN-LSTM con una ventana temporal de 30 días y muestras de tamaño 64 en su capa. Este modelo logra un coeficiente de determinación (R^2) del 98 %, lo que resulta en un modelo de gran precisión.

Abstract

Accurate prediction of short-term stock prices plays a crucial role in financial markets, as it provides investors with the ability to make informed decisions and maximize their returns. While it was previously believed that stock price time series were unpredictable, recent research has revealed the existence of patterns and trends that allow predictions to be made with some accuracy. In this context, deep learning techniques have emerged as a powerful tool, as they are able to capture complex patterns and adapt to changes in the data. In order to take advantage of these benefits, this thesis focuses on a comparison of strategies based on deep learning to predict in the short term the closing price of the shares of Acciona, a company of great relevance in the IBEX 35 stock index. Through the exhaustive analysis and evaluation of different techniques and models, we seek to improve the prediction capacity.

For this purpose, the download and preprocessing of the historical data of the closing prices of Acciona's shares from 2015 to 2023 is carried out. During this process, data normalization is carried out and the absence of missing values is verified to ensure the integrity of the time series. Subsequently, the division of the time series into training and test sets is performed, assigning 60% of the data for the training set and 40% for the test set. In addition, the 'early stopping' strategy is implemented in order to improve model performance and avoid overfitting. Within the implementation, the use of different window sizes is explored, including 30, 60 and 90-day options, with different sample size alternatives, such as 16, 32 and 64. These variations in window and sample sizes allow us to evaluate the influence of these parameters on the predictive ability of the model and to determine the optimal combination for accurate results. After implementing and comparing different techniques of neural network architectures, it is concluded that the model that best fits the historical data of Acciona's stock prices is the one based on a CNN-LSTM architecture with a 30-day time window and sample size 64 in its layer. This model achieves a coefficient of determination (R^2) of 98%, resulting in a highly accurate model.

Índice general

1. Introducción	1
1.1. Motivación	1
1.2. Objetivos del Trabajo de Grado	3
1.2.1. Objetivo General	3
1.2.2. Objetivos específicos	3
1.3. Organización de la memoria	3
2. Predicción de series temporales en el mercado de valores usando aprendizaje profundo: una revisión de la literatura	5
3. Estrategias de predicción de series de tiempo usando Deep Learning	14
3.1. Redes Neuronales Recurrentes	14
3.2. Long Short-Term Memory	15
3.3. Bidirectional Long Short-Term Memory	17
3.4. Stacked Long Short-Term Memory	19
3.5. Convolutional Neural Network	20
3.6. Convolutional Neural Network Long Short-Term Memory	21
3.7. Métricas de evaluación	23
4. Predicción a corto plazo de acciones usando Deep Learning	25
4.1. Esquema Metodológico	25
4.2. Adquisición de datos y pre-procesamiento	26
4.3. Análisis Descriptivo	28
4.4. Implementación de modelos	34
4.4.1. Modelo LSTM Básico	36
4.4.2. Modelo Stacked LSTM	37
4.4.3. Modelo Bi-LSTM	37
4.4.4. Modelo CNN-LSTM	38
4.5. Evaluación de los resultados	38
4.5.1. Resultados Modelo LSTM	39
4.5.2. Resultados Modelo Stacked LSTM	40

4.5.3. Resultados Modelo BiLSTM	42
4.5.4. Resultados Modelo CNN-LSTM	42
4.5.5. Comparativa general de los modelos implementados	44
5. Conclusiones y trabajo futuro	46
Bibliografía	51

Índice de figuras

3.1. Esquema comparativo de la arquitectura de Redes Neuronales Estándar, Recurrentes y LSTM, donde X es la entrada e Y la salida correspondiente Elaboración Propia	17
3.2. Representación gráfica de la arquitectura LSTM bilateral Elaboración propia. Adaptada de (O.Nawajana, 2020)	18
3.3. Estructura jerárquica de un Stacked LSTM Elaboración Propia	20
3.4. Ejemplo de una CNN con 6 capas Elaboración Propia. Adaptada de (Delteil, 2023) .	21
3.5. Arquitectura CNN-LSTM Elaboración Propia	22
4.1. Enfoque metodológico utilizado para llevar a cabo la predicción del valor de cierre de una acción	27
4.2. Gráfico de Vela de los precios de apertura y cierre de las acciones de Acciona	29
4.3. Distribución de las variables obtenidas en el data-set	31
4.4. Precio de cierre a lo largo del tiempo	32
4.5. Precio de cierre a lo largo del tiempo	33
4.6. Precio de cierre a lo largo del tiempo	34
4.7. Representación del resultado del modelo LSTM 30-16 Elaboración propia . . .	40
4.8. Representación del resultado del modelo Stacked LSTM 90-32-32 Elaboración propia	42
4.9. Representación del resultado del modelo BiLSTM 60-32 Elaboración propia .	43
4.10. Representación del resultado del modelo CNN LSTM 30-64 Elaboración propia	45

Índice de tablas

2.1. Comparativa de artículos que abordan las diferentes técnicas la predicción de la cotización de acciones a corto plazo	13
4.1. Definición de las variables del dataset	28
4.2. Estadísticas de las variables del conjunto de datos analizado	29
4.3. Resultado de las variaciones en los parámetros para el modelo LSTM- Un paso adelante	40
4.4. Resultado de las alteraciones al modelo Stacked LSTM	41
4.5. Resultado de las diversas configuraciones del modelo BiLSTM	43
4.6. Resultado de las alteraciones al modelo CNN-LSTM	44
4.7. Conclusión de resultados	45

Acrónimos

<i>LSTM</i>	Long-Short Term Memory
<i>BiLSTM</i>	Bidirectional Long-Short Term Memory
<i>GRU</i>	Gated Recurrent Unit
<i>ANN</i>	Artificial Neural Networks
<i>MOCWT</i>	Multi Optic Combined Wavelet Transformation
<i>ICA</i>	Análisis de Componentes Independientes
<i>EEB</i>	Bolsa de Bombay
<i>ARIMA</i>	Auto Regressive Integrated Moving Average
<i>ETS</i>	Error Trend Seasonality
<i>RNN</i>	Recurrent Neural Network
<i>MAPE</i>	Mean Absolute Percentage Error
<i>MAE</i>	Mean Absolute Error
<i>RMSE</i>	Root Mean Square Error
<i>MSE</i>	Mean Squared Error
<i>TLBO</i>	Teaching Learning based Optimization
<i>RNA</i>	Redes Neuronales Artificiales
<i>ReLU</i>	Unidad Lineal Rectificada
<i>SGD</i>	Descenso de Gradiente Estocástico
<i>CNN</i>	Convolutional Neural Network
<i>R²</i>	Coefficiente de Determinación
<i>MLP</i>	Multilayer Perceptron
<i>SACLSTM</i>	Long-Short Term Memory convolucional de secuencias de acciones
<i>BiSLSTM</i>	Special bidirectional long-term memory
<i>SVM</i>	Support Vector Machine
<i>NLP</i>	Lenguaje Natural
<i>IQR</i>	Rango Intercuartil

Capítulo 1

Introducción

1.1. Motivación

Los índices de precios de las acciones son muy significativos entre los mercados financieros mundiales como criterio principal para evaluar la función del mercado de valores. Estas cotizaciones se obtienen a partir de la acumulación de los movimientos de los precios de las acciones a corto plazo para ciertas empresas individuales o para un gran conjunto de empresas presentes en el mercado bursátil (J.-J. Wang, Wang, Zhang, y Guo, 2012). En este contexto, debido a que los mercados financieros presentan un alto grado de competencia entre los participantes (Parot, Michell, y Kristjanpoller, 2019), el estudio de los modelos de predicción de los índices de precios de las acciones no sólo aporta un gran beneficio económico al accionista, sino también un valor de conocimiento necesario para los inversores, los cuales podrán contar con información de utilidad en la selección de una cartera apropiada y en la provisión de oportunidades de inversión internacional.

En los últimos años, debido al rápido desarrollo económico, son cada vez más las personas que comienzan a invertir en el mercado de valores. Es por ello, que la comprensión de los patrones que caracterizan la dinámica de las series de tiempo son temas de investigación en ámbitos tanto académicos como financieros. Una de las principales preocupaciones de los inversores es obtener una visión real del futuro y de la tendencia predominante en los datos, sobre todo teniendo en cuenta que las fluctuaciones en datos financieros dependen de una variedad de factores correlacionados, que cambian constantemente (Jansen, 2020). En este sentido, la importancia de obtener una predicción fiable de los índices bursátiles permite mejorar el desarrollo de actividades como:

1. **Decisiones de inversión:** Los inversores utilizan las predicciones de los índices bursátiles para tomar decisiones de inversión informadas, como la compra o venta de acciones.
2. **Análisis del mercado:** Las predicciones ayudan a los analistas a comprender las ten-

dencias del mercado y a tomar decisiones informadas sobre el futuro del mercado.

3. **Gestión del riesgo:** Las predicciones pueden ayudar a gestionar el riesgo de inversión al proporcionar información sobre los posibles movimientos del mercado.
4. **Diversificación de la cartera:** Las predicciones precisas pueden utilizarse para diversificar las carteras, equilibrando las inversiones en diferentes industrias y sectores.

En el contexto de este campo de investigación, durante mucho tiempo se ha creído que las series de tiempo de cotizaciones bursátiles no son predecibles debido a la complejidad y la imprevisibilidad del mercado. Si bien es cierto que el mercado de valores está sujeto a una gran cantidad de factores externos e internos que pueden influir en los movimientos de los precios de las acciones, como la economía global, la política, la inflación, entre otros, también es cierto que estos factores no son totalmente impredecibles. Es así como algunos estudios recientes han demostrado que la identificación de patrones y anomalías en los movimientos de los precios de las acciones, como las fluctuaciones repentinas y las tendencias a largo plazo, podrían permitir la predicción de los índices de precios de las acciones con cierto grado de precisión. De hecho, en los últimos años, se ha observado un cambio en la forma en que los inversores abordan la toma de decisiones de inversión en el mercado de valores. En lugar de depender únicamente de la intuición y la experiencia, los inversores han comenzado a utilizar datos históricos de precios de las acciones para identificar patrones y tendencias en los movimientos del mercado. Este cambio en la estrategia de inversión sugiere que la información contenida en los datos históricos de la serie temporal puede ser una herramienta valiosa para predecir el comportamiento futuro del mercado (Murphy John, 1999).

De esta manera, se ha desarrollado una gran variedad de metodologías a lo largo de la historia para el análisis y la predicción de series temporales de cotizaciones bursátiles, teniendo en cuenta la información temporal que estas series brindan. Entre las técnicas más comunes utilizadas en este campo se encuentran los modelos autorregresivos integrados de media móvil (ARIMA), los modelos de suavizamiento exponencial, y los modelos de regresión (Ballman, 1998; Song, 2017). A pesar de la efectividad de estas técnicas, su limitación radica en su capacidad para modelar datos complejos y no lineales. Es por ello que durante la última década, el desarrollo de técnicas de inteligencia artificial, como el aprendizaje profundo, ha permitido la creación de modelos más precisos y complejos para la predicción de series temporales de cotizaciones bursátiles. La capacidad de estos modelos para capturar patrones no lineales y su habilidad para adaptarse a los cambios en los datos los convierte en una herramienta poderosa para la predicción del comportamiento futuro del mercado bursátil.

El uso de estas técnicas ofrece varias ventajas significativas, como la mejora de la precisión de las predicciones, la reducción de errores de predicción y la capacidad para identificar patrones complejos en los datos (Huang, Chai, y Cho, 2020). Además, estas técnicas pueden ser aplicadas en tiempo real para actualizar las predicciones a medida que se reciben nuevos

datos, lo que permite una gestión más efectiva del riesgo y una toma de decisiones de inversión más informada. En consecuencia, las empresas y los inversores pueden aprovechar estas herramientas para optimizar su rendimiento y aumentar su capacidad para tomar decisiones más informadas y estratégicas en el mercado bursátil. Teniendo en cuenta los desafíos que enfrenta la predicción de las cotizaciones bursátiles y la relevancia de una precisión cada vez mayor en la toma de decisiones de inversión, resulta evidente la necesidad de explorar nuevas técnicas y algoritmos capaces de abordar estos retos. Este es el contexto principal en el que se enmarca la línea de investigación del presente trabajo de fin de grado.

1.2. Objetivos del Trabajo de Grado

1.2.1. Objetivo General

El objetivo principal de este trabajo de fin de grado es analizar los modelos predictivos basados en deep learning para estimar el valor futuro, a corto plazo, de las acciones de una empresa que cotiza en bolsa. Para ello, se realizará un análisis comparativo de las técnicas presentes en la literatura y se concluirá con una propuesta de modelo que presente las mejores métricas de desempeño.

1.2.2. Objetivos específicos

1. Contextualizar la importancia de las técnicas de predicción basadas en deep learning para estimar los valores futuros de la cotización de acciones en bolsa.
2. Desarrollar un modelo de predicción a corto plazo basado en la implementación de diferentes métodos de deep learning, teniendo como base los datos históricos de cotización en bolsa de una compañía.
3. Evaluar el rendimiento de las diferentes metodologías y concluir sobre la estrategia que mejor desempeño presenta en la predicción del precio de cotizaciones en bolsa.

1.3. Organización de la memoria

La memoria de este trabajo se divide en cinco capítulos. En el Capítulo 2, se presenta una revisión exhaustiva de la literatura sobre las investigaciones dedicadas a la predicción a corto plazo de series de tiempo en cotizaciones del mercado de valores, utilizando diferentes algoritmos de Deep Learning. A continuación, en el Capítulo 3, se describe en detalle el marco teórico de las técnicas a utilizar, así como la metodología de procesamiento de cada estrategia. En el Capítulo 4, se presenta la implementación de los modelos y los resultados

obtenidos en el estudio. Por último, en el Capítulo 5, se discuten las principales conclusiones y se describe el trabajo futuro asociado a esta línea de investigación.

Capítulo 2

Predicción de series temporales en el mercado de valores usando aprendizaje profundo: una revisión de la literatura

La predicción del precio de las acciones a corto plazo es fundamental para que los inversores y traders tomen decisiones informadas en los mercados financieros. Estos modelos se enfocan en prever las futuras fluctuaciones de los precios de las acciones, lo que puede proporcionar oportunidades para maximizar ganancias o evitar pérdidas. En este contexto, el desarrollo de la inteligencia artificial ha llevado a un creciente uso de los métodos de aprendizaje profundo en la predicción del precio de las acciones. El aprendizaje profundo es especialmente relevante en este campo debido a su capacidad para procesar eficientemente grandes volúmenes de datos, lo cual es esencial en el análisis financiero, donde se dispone de una gran cantidad de información en tiempo real. En comparación con los métodos estadísticos tradicionales, el aprendizaje profundo ha demostrado un rendimiento notablemente superior en términos de precisión y capacidad de modelado (Liang, Ge, Sun, He, y Chen, 2019). En este contexto, diversos estudios basados en este enfoque se han llevado a cabo, proponiendo técnicas y modelos innovadores con el objetivo de mejorar la precisión en la predicción de precios a corto plazo.

En este punto, es importante destacar que el proceso de toma de decisiones precisas en los mercados financieros involucra el uso de diferentes enfoques de análisis. Según (Bagheri, Peyhani, y Akbari, 2014), los inversores se basan en dos tipos principales de análisis: el análisis fundamental y el análisis técnico. El análisis fundamental se centra en indicadores industriales, comerciales y económicos globales para evaluar el valor intrínseco de las acciones. Por otro lado, el análisis técnico se basa en el estudio de los precios históricos, asumiendo que los comportamientos pasados influyen en la evolución futura de los rendimientos bursátiles (Patel, Shah, Thakkar, y Kotecha, 2015). En general, puede decirse que los inversores utilizan el análisis técnico para explorar el comportamiento del mercado y anticipar las

tendencias futuras. A nivel micro, los inversores pueden beneficiarse al aplicar el análisis técnico en la toma de decisiones de inversión a corto plazo.

A lo largo de los años, varios autores han sostenido argumentos en favor de uno u otro enfoque (Bülow, 2017; Masry et al., 2017). Por esta razón, en el estudio de (Kofi, Adekoya, y Weyori, 2020), se propuso revisar y comparar la precisión de los estudios realizados en años anteriores. La investigación realizó una revisión sistemática y crítica de aproximadamente 122 trabajos publicados en el período de 11 años (2007-2018) en el campo de la predicción del mercado de valores utilizando estrategias de Machine Learning. A partir de estos informes, se identificaron diversas técnicas, que se agruparon en tres categorías: análisis técnico, análisis fundamental y análisis combinado. Los resultados revelaron que el 66 % de los documentos revisados se basaban en análisis técnico, mientras que el 23 % y el 11 % se basaban en análisis fundamental y análisis combinado, respectivamente. En cuanto al número de fuentes de datos utilizadas, se observó que el 89.34 % de los documentos examinados empleaba una sola fuente, mientras que el 8.2 % y el 2.46 % utilizaban dos y tres fuentes, respectivamente. En términos de algoritmos de aprendizaje automático, se encontró que las máquinas de vectores de soporte y las redes neuronales artificiales eran los más utilizados para la predicción bursátil.

En el contexto del análisis técnico y la predicción de acciones a corto plazo, se han propuesto diversas estrategias para abordar esta tarea compleja, que implica analizar un volumen significativo de información en tiempo real y de naturaleza no lineal. A lo largo del tiempo, se han desarrollado modelos predictivos que demuestran que una de las técnicas de Machine Learning más precisas son las redes neuronales, principalmente aquellas arquitecturas de aprendizaje profundo, gracias a su capacidad para manejar grandes volúmenes de datos con relaciones complejas, y de identificar patrones. En el ámbito de las redes neuronales, existen numerosos tipos y variaciones de arquitecturas que se han empleado en la predicción del mercado de valores. Estos enfoques difieren en su estructura y funcionamiento, y cada uno ofrece ventajas y desventajas en términos de su capacidad para capturar patrones y tendencias en los datos financieros.

En esta línea de investigación, (Sethia y Raut, 2019) presentan un modelo desarrollado para la predicción de los precios de las acciones a cinco días ($t + 5$), con el objetivo de proporcionar una estrategia diaria de compra/venta para el índice Standard's and Poor's 500 (S&P 500). En dicho estudio, se llevó a cabo una comparación de rendimiento entre diferentes modelos, incluyendo LSTM (Long Short-Term Memory), GRU (Gated Recurrent Unit), ANN (Artificial Neural Networks) y SVM (Support Vector Machine), con el propósito de determinar el modelo óptimo para esta tarea específica. El conjunto de datos utilizado para el entrenamiento y la predicción abarcó un periodo de 17 años, desde 2000 hasta 2017. Se generaron 50 atributos basados en indicadores técnicos, los cuales se añadieron a los datos de apertura, máximos, mínimos, cierre y volumen del índice S&P 500. Cada valor de atributo se transformó en una puntuación estándar relativa, seguida de un escalado minimax y

una reducción de la dimensionalidad mediante la técnica de ICA (Análisis de Componentes Independientes). Posteriormente, se compararon los resultados obtenidos por los distintos modelos utilizando métricas definidas por los investigadores. Entre estas métricas se encontraban el coeficiente de optimismo y pesimismo, así como el coeficiente de rentabilidad. Los resultados revelaron que el modelo LSTM superó a los demás en términos de rentabilidad, alcanzando un rendimiento un 400 % superior a la estrategia de mantener y esperar. Además, el modelo LSTM obtuvo una alta puntuación de R2 (Coeficiente de Determinación), con un valor de 0,9486.

Un estudio adicional de interés en el área es el llevado a cabo por (Liang et al., 2019) en respuesta a la problemática del ruido presente en los datos financieros. Para abordar este desafío, los autores de este estudio adhieren el método wavelet threshold-denoising a una red neuronal LSTM, enfocándose en el preprocesamiento de datos mediante el uso de un modelo de umbral (suave/duro), lo cual permite mitigar de manera evidente el ruido presente en la señal financiera. Como una contribución novedosa, los autores proponen un nuevo método denominado MOCWT (Multi Optic Combined Wavelet Transformation), que introduce una función de umbral-denoising mejorada con el objetivo de reducir la distorsión en la reconstrucción de la señal. Los resultados experimentales obtenidos revelan claramente que la MOCWT propuesta supera a los métodos tradicionales en términos de precisión de la predicción, lo cual confirma su eficacia en el contexto de la predicción de precios de acciones. En el estudio realizado, se utilizaron como datos fuente los precios de apertura del índice S&P 500 durante un periodo de dieciocho años. Los resultados experimentales destacan el rendimiento mejorado del método de procesamiento de umbrales duros, donde la pérdida mínima alcanzó un valor de 2,453.

Otro estudio que respalda la eficacia de la arquitectura de redes LSTM es el propuesto por (Bhanja y Das, 2019). En este estudio, los autores reconocen que a lo largo de los años se han realizado diversos estudios que presentan alternativas al modelo LSTM estándar, y en su investigación se centra en el desarrollo de un modelo con arquitectura LSTM apiladas (stacked LSTM) con el objetivo de demostrar su eficiencia mejorada. En su trabajo, los autores utilizan datos históricos de series temporales de la EEB (Bolsa de Bombay) recopilados desde el 1 de enero de 2014 hasta el 31 de diciembre de 2018, obtenidos a partir del sitio web de Yahoo Finanzas. Los resultados obtenidos en su investigación demuestran claramente que una arquitectura LSTM con múltiples unidades supera a la versión estándar. Esta mejora se debe a varias razones fundamentales. En primer lugar, la arquitectura stacked LSTM actúa como un regulador natural para controlar el sobreajuste del modelo, lo que contribuye a una mayor generalización y precisión en las predicciones. Además, al agregar unidades LSTM adicionales, se aumenta la capacidad del modelo para capturar relaciones complejas y extraer características más sofisticadas de los datos financieros. Esto implica que el modelo puede representar de manera más precisa los patrones y las tendencias subyacentes en los datos. Además, la arquitectura stacked LSTM posee una capacidad inherente para aprender repre-

sentaciones jerárquicas. Cada unidad puede aprender características a diferentes niveles de abstracción, lo que permite capturar patrones y relaciones más sofisticadas en los datos financieros. Esta capacidad de aprendizaje de representaciones jerárquicas ayuda al modelo a capturar mejor la complejidad de los movimientos del mercado y a generar predicciones más precisas.

En esta línea de investigación, (Lu, Li, Li, Sun, y Wang, 2020) consideraban la LSTM como una técnica óptima para la predicción de acciones a corto plazo. Sin embargo, buscaban demostrar que al implementarla junto con las capacidades de pooling para extraer características espaciales de los datos de entrada de un CNN (Convolutional Neural Network), su eficiencia mejoraría significativamente. Para comprobar la precisión del modelo CNN-LSTM propuesto, se llevó a cabo una comparación con otros métodos, como MLP (Multilayer Perceptron), CNN, RNN (Recurrent Neural Network), LSTM y CNN-RNN, utilizando el mismo conjunto de datos de entrenamiento y validación en un entorno operativo similar. Los datos utilizados en este estudio consistieron en información sobre las acciones del índice de Shanghai Composite desde el 1 de julio de 1991 hasta el 31 de agosto de 2020. Se analizaron variables como el precio de apertura, el precio máximo, el precio mínimo, el precio de cierre, el volumen, las subidas y bajadas, entre otros. Utilizando esta información, se realizó la predicción del precio de cierre del día siguiente, dividiendo los 6,627 días en un conjunto de entrenamiento y 500 días en un conjunto de prueba. Al comparar los resultados obtenidos por el modelo CNN-LSTM con los del MLP, CNN, RNN, LSTM y CNN-RNN, se demostró que el modelo CNN-LSTM exhibe una alta precisión en la predicción y es especialmente adecuado para predecir el precio de las acciones a corto plazo.

En el ámbito financiero, se reconoce que la demanda de las acciones es un factor esencial en la determinación de sus precios. Existe una relación directa entre la demanda y el precio de las acciones, siendo que un aumento en la demanda conlleva a un incremento en su precio. Consciente de esta relación, (Abbasimehr, Shabani, y Yousefi, 2020) realizan una revisión de técnicas previamente utilizadas (Bhanja y Das, 2019; Liang et al., 2019), específicamente el método LSTM y Stacked LSTM. No obstante, su enfoque se centra en la predicción de la demanda, con el objetivo de justificar los precios de cada activo. Para alcanzar este objetivo, los autores aplican el método LSTM multicapa a los datos de demanda de una empresa de muebles. Su investigación se centra en encontrar los hiperparámetros óptimos de la red LSTM, aprovechando la capacidad inherente de esta arquitectura para capturar patrones no lineales. Con el fin de evaluar la efectividad del modelo propuesto, se comparan los resultados obtenidos con otras técnicas utilizadas en el análisis financiero, tanto métodos estadísticos tradicionales como técnicas de inteligencia computacional, como ARIMA (Auto Regressive Integrated Moving Average), ETS (Error Trend Seasonality), ANN y SVM. Los resultados experimentales demuestran que el modelo propuesto supera sustancialmente a las alternativas consideradas en términos de precisión y rendimiento. Esto destaca la eficacia de la arquitectura Stacked LSTM en la predicción de la demanda de acciones y respalda su

utilidad en el análisis financiero.

Con el objetivo de ampliar el alcance de la investigación, (Nabipour, Nayyeri, Jabani, Mosavi, y Salwana, 2020) llevaron a cabo una validación exhaustiva del uso de estrategias basadas en redes neuronales, ampliando el espectro de técnicas de comparación al incluir varias otras basadas en Machine Learning. Para realizar las evaluaciones experimentales, se seleccionaron cuatro grupos de datos del archivo en línea de la Tehran Securities Exchange Technology Management Co., que incluían sectores financieros diversificados, petróleo, minerales no metálicos y metales básicos. Se recopilaron datos históricos para estos grupos durante un período de 10 años, desde 2009 hasta 2019. Con el propósito de realizar pronósticos de valores con diferentes horizontes de tiempo, se generaron predicciones para 1, 2, 5, 10, 15, 20 y 30 días con antelación. Para llevar a cabo este análisis comparativo, se emplearon seis modelos basados en árboles, como Decision Tree, Bagging, Random Forest, Adaboost, Gradient Boosting y XGBoost, junto con tres algoritmos basados en redes neuronales: ANN, RNN y LSTM. Los resultados obtenidos se evaluaron utilizando métricas tradicionales en la predicción de series de tiempo como MAPE (Mean Absolute Percentage Error), MAE (Mean Absolute Error), RMSE (Root Mean Square Error) y MSE (Mean Squared Error). Entre todos los algoritmos utilizados en este estudio, se observó que el modelo LSTM demostró los resultados más precisos, mostrando una mayor capacidad de ajuste del modelo en comparación con los demás. En el caso de los modelos basados en árboles, se observó métricas de desempeño similares entre Adaboost, Gradient Boosting y XGBoost. Sin embargo, en general, se observó un aumento en los valores de error al crear modelos de predicción para un mayor número de días de antelación. Por ejemplo, los valores de MAPE para XGBoost fueron de 0,88, 1,14, 1,45, 1,77, 2,03, 2,30 y 2,48, para cada uno de los horizontes de predicción expuestos anteriormente.

Basándose en la investigación previa sobre el uso de algoritmos basados en unidades LSTM para la predicción de acciones a corto plazo, el estudio realizado por (J. M.-T. Wu, Li, Herencsar, Vo, y Lin, 2021), se basa en los trabajos desarrollados por (Lu et al., 2020) y propone un modelo denominado SACLSTM (Long-Short Term Memory convolucional de secuencias de acciones), el cual combina las ventajas de las redes convolucionales y las redes LSTM para lograr una predicción más precisa del precio de las acciones. En este artículo, se introduce una nueva estructura que utiliza una matriz secuencial de datos históricos y sus indicadores principales, como entrada de la arquitectura CNN. A través de la capa convolucional, se extraen vectores de características específicos que posteriormente son analizados con unidades LSTM. El modelo SACLSTM se evalúa utilizando una serie temporal de diez años en EE. UU. y Taiwán como datos experimentales. La muestra utilizada comprende datos históricos, futuros y otros indicadores de acciones de empresas del mercado estadounidense (Apple, IBM, Microsoft, Meta y Amazon) y del mercado taiwanés (Compagnie des Alpes, Formosa Plastics, Asustek Computer, MediaTek y Largan Precision). Los resultados obtenidos demuestran que la arquitectura SACLSTM tiene un rendimiento sobresaliente en compa-

ración con los modelos CNN y LSTM, individualmente. En particular, el modelo SACLSTM alcanza una precisión del 91.1 % en la predicción a 1 día, superando el rendimiento del 78 % y 86 % de los modelos CNN y LSTM, respectivamente. Estos resultados respaldan la efectividad del enfoque propuesto y destacan su capacidad para mejorar la precisión de la predicción de precios de acciones.

En relación al trabajo previo realizado por (J. M.-T. Wu et al., 2021), (H. Wang et al., 2021) se interesó en investigar si se podía mejorar la precisión de la predicción mediante la creación de una arquitectura bidireccional. En su estudio, proponen un modelo llamado CNN-BiSLSTM (Special bidirectional long-short term memory) para predecir el precio de cierre de las acciones. El modelo CNN-BiSLSTM utiliza una variante de la red BiSLSTM, la cual incorpora la función $1 - \tanh(x)$ en la compuerta de salida, mejorando la predicción del precio de las acciones. El modelo extrae características avanzadas que influyen en el precio de las acciones mediante una red CNN y, posteriormente, realiza la predicción del precio de cierre utilizando la red BiSLSTM. Para evaluar la eficacia del modelo propuesto, se utilizan datos históricos del índice de componentes de Shenzhen desde el 1 de julio de 1991 hasta el 30 de octubre de 2020. El desempeño de CNN-BiSLSTM se compara con varios otros modelos, como MLP, RNN, LSTM, BiLSTM (Bidirectional Long-Short Term Memory), CNN-LSTM y CNN-BiLSTM. Los resultados experimentales demuestran que, en comparación con LSTM, la red BiLSTM logra reducir el MAE en aproximadamente un 5,89 % y el RMSE en aproximadamente un 5,24 %. Esto indica que el modelo CNN-BiSLSTM puede predecir con precisión el precio de cierre del índice de componentes de Shenzhen para el siguiente día de negociación.

Recientemente, (Swathi, Kasiviswanath, y Rao, 2022) propusieron una solución alternativa mediante la implementación de la técnica LSTM, incorporando opiniones públicas de las personas, quienes también pueden ser posibles inversores. En el actual mercado bursátil, las opiniones positivas y negativas juegan un papel crucial como indicadores de las futuras cotizaciones. Por lo tanto, el análisis del sentimiento en los datos de las redes sociales relacionados con los precios de las acciones puede ser de gran ayuda para predecir de manera efectiva los futuros precios de las acciones. Con esta motivación, la investigación presentó un nuevo modelo de optimización basado en la técnica de enseñanza y aprendizaje: TLBO (Teaching Learning based Optimization) junto con la implementación de una estrategia LSTM, basándose en el análisis del sentimiento para la predicción de los precios de las acciones utilizando datos de Twitter. Además, se utilizó el modelo LSTM para clasificar los tweets en sentimientos positivos y negativos relacionados con los precios de las acciones, lo que permitió investigar la correlación entre los tweets y la naturaleza de las cotizaciones bursátiles. Para mejorar la precisión predictiva del modelo LSTM se utilizó el modelo TLBO para ajustar de manera óptima la unidad de salida del modelo LSTM. Se realizaron experimentos con datos de Twitter para evaluar el rendimiento predictivo del modelo TLBO-LSTM. Los resultados experimentales obtenidos por el modelo TLBO-LSTM demostraron resultados prometedores

en comparación con otras técnicas del estado del arte. El modelo TLBO-LSTM logró un resultado superior, con una precisión máxima del 95,33 %, un recall del 85,28 % y un F-score del 90 %. Al obtener una precisión del 94,73 %, el modelo TLBO-LSTM superó a las demás técnicas evaluadas.

En conclusión, el presente capítulo ha abordado el estudio y comparación de diversas técnicas basadas en deep learning, utilizadas en la predicción de precios de acciones en el ámbito financiero. Sin embargo, es importante destacar que realizar una comparación directa entre estas técnicas resulta desafiante debido a las diferencias en los conjuntos de datos utilizados, las dinámicas del mercado y los horizontes temporales considerados en cada estudio. Estas variaciones pueden afectar significativamente los resultados y dificultan la obtención de conclusiones sólidas sobre la superioridad de una técnica sobre otra. Por lo tanto, en este trabajo de fin de grado se busca llenar esta brecha, realizando una comparación entre las diferentes técnicas mencionadas con un conjunto de datos enmarcado en el contexto del mercado español. Al enfocarse en un mercado específico, se espera reducir las discrepancias y proporcionar una visión más precisa de la eficacia y las limitaciones de cada enfoque.

Tras realizar la revisión de la literatura existente en este campo de aplicación, se presenta en la Tabla 2.1 un resumen de las técnicas implementadas hasta el momento. Esta tabla sintetiza de manera concisa las diferentes metodologías y enfoques utilizados en los estudios previos, brindando una visión general de las diversas herramientas empleadas para la predicción de precios de acciones en el ámbito financiero.

Como puede evidenciarse en la Tabla 2.1, la mayoría de las soluciones propuestas para la predicción a corto plazo del precio de las acciones, basadas en aprendizaje profundo, involucran arquitecturas simples o híbridas basadas en unidades LSTM. Sin embargo, se ha observado que las arquitecturas CNN-LSTM han demostrado ligeramente mejores resultados en términos de precisión de predicción. Estas arquitecturas combinan las capacidades de extracción de características espaciales de las CNN con la capacidad de análisis de relaciones temporales de las LSTM. En la mayoría de los estudios revisados, se ha observado que las variables de entrada analizadas se centran en aspectos clave como el precio de apertura, el precio de cierre, el precio mínimo y máximo del día, así como el volumen de negociación. Entre todas estas variables, se destaca que el precio de cierre es la más utilizada como variable de entrada en los modelos propuestos. La elección del precio de cierre como variable de entrada se basa en que este factor refleja el valor final al que se ha negociado una acción en un día determinado, lo que proporciona una medida consolidada del comportamiento del mercado y la evaluación de los inversores al final de la jornada. Esto lo convierte en un indicador clave para comprender la tendencia y el rendimiento de una acción.

Referencia	Objetivo	VARIABLES DE ENTRADA	Etapa temporal	Técnicas utilizadas	Principales resultados
(Sethia y Raut, 2019)	Desarrollar un modelo óptimo para predecir los precios de las acciones para el día $t+5$ y generar una estrategia diaria de compra / venta para el índice Standard & Poor's 500	Datos de apertura, máximos, mínimos, cierre y volumen del índice S&P 500	Durante 17 años, desde el año 2000 al 2017	LSTM, ICA, GRU y ANN	El modelo LSTM superó a los demás modelos con una rentabilidad un 400 % mayor y una puntuación R^2 de 0,9486.
(Liang et al., 2019)	Mejorar la precisión de la predicción de precios de acciones mediante el uso del método de umbral-noising de wavalet	Precio de cierre del índice S&P500	18 años	LSTM con transformación wavalet utilizando medidas de evaluación como MSE o R2	El rendimiento del método de procesamiento de umbral duro mejorado mejora notablemente. La pérdida mínima alcanza 2.453.
(Bhanja y Das, 2019)	Aumentar la precisión y estabilidad de la predicción del mercado de valores mediante un modelo híbrido basado en redes neuronales	Precio de cierre	Desde el 1 de enero de 2014 hasta el 31 de diciembre de 2018	LSTM y Stacked LSTM	Una arquitectura LSTM con más de una capa es más eficiente que la estándar demostrado con un MSE, MAE y MAPE de 0.00152, 0.01083 y 0.3048 respectivamente.
(Lu et al., 2020)	Mejorar la precisión en la predicción a corto plazo de los precios de las acciones mediante un enfoque basado en el aprendizaje profundo	Precio de apertura, precio máximo, precio mínimo, precio de cierre, volumen, subidas y bajadas, y cambio.	Desde el 1 de julio de 1991 hasta el 31 de agosto de 2020	CNN-LSTM, MLP, CNN, RNN, LSTM y CNN-RNN	Los resultados muestran que el rendimiento de CNN-LSTM es el mejor de los seis métodos. En términos de precisión de la previsión, el MAE es de 27,564 y el RMSE de 39,688, que es el menor de los seis modelos de previsión. Además, el R2 de es de 0,9646, que mejora en un 2,2 %, 0,6 %, 0,5 % y 0,2 %, respectivamente, en comparación con los otros cuatro métodos.
(Abbasimehr et al., 2020)	Mejorar la predicción del precio de las acciones capturando patrones y tendencias de los datos históricos de demanda a través de un modelo LSTM	Datos históricos de la demanda de muebles	132 meses desde el 2007 al 2017	LSTM, ETS, ARIMA, ANN, SVM, RNN y KNN	Comparando todas las técnicas, destaca la eficiencia de la técnica LSTM (mejora si es multicapa) para modelar patrones de una serie temporal.
(Nabipour et al., 2020)	Estudiar las diferentes arquitecturas y configuraciones de modelos de aprendizaje profundo para proponer una técnica con una mejor precisión de predicción del mercado de valores	Precio de apertura, cierre, máximo y mínimo de 4 grupos bursátiles (finanzas diversificadas, petróleo, minerales no metálicos y metales básicos)	Durante 10 años, entre 2009 y 2019	Seis modelos basados en árboles, a saber, Decision Tree, Bagging, Random Forest, Adaboost, Gradient Boosting y XG-Boost, y también ANN, RNN y LSTM	El modelo LSTM se mostró más efectivo para predecir los precios de las acciones con un MAPE de 0.60.

Referencia	Objetivo	VARIABLES DE ENTRADA	Etapa temporal	Técnicas utilizadas	Principales resultados
(J. M.-T. Wu et al., 2021)	Mejorar la precisión de predicción del precio de acciones a 1,3 y 7 días construyendo una matriz secuencial de datos históricos	Precio actual, el precio máximo, el precio de apertura, el precio mínimo, el volumen y las subidas y bajadas, futuros y opciones de las acciones de AAPL, IBM, MSFT, FB y AMZN, CDA, CFO, DJO, DVO y IJO	120 días	CNN, LSTM y SACLSTM	Los resultados muestran que la arquitectura SACLSTM tiene el mejor rendimiento con una precisión del 91.1 % en la predicción a 1 día, en comparación del 78 % y 86 % del CNN y LSTM, respectivamente.
(H. Wang et al., 2021)	Desarrollar un modelo de arquitectura bidireccional para mejorar la precisión de predicción de las acciones de Shenzhen	Precio de apertura, de cierre, máximo, mínimo, volumen y cambio	Desde el 1 de julio de 1991 hasta el 30 de octubre de 2020	CNN-BiLSTM, MLP, la RNN, LSTM, BiLSTM, CNN-LSTM y la CNN-BiLSTM	Los resultados experimentales muestran que, en comparación con LSTM, el MAE de BiLSTM se reduce en aproximadamente 5,89 y el RMSE en aproximadamente 5,24.
(Swathi et al., 2022)	Mejorar la precisión de la predicción de acciones combinando el análisis de sentimientos en Twitter con un modelo LSTM y su repercusión en los precios	n.a	Análisis de sentimiento de Twitter y LSTM con métricas de precisión como MAE, MSE y R2	El modelo TLBO-LSTM produjo un resultado superior, con una precisión máxima del 95,33 %, un recall del 85,28 % y un F-score del 90 %. Al lograr una mayor precisión del 94,73 %, el modelo TLBO-LSTM superó a las demás técnicas.	

Tabla 2.1: Comparativa de artículos que abordan las diferentes técnicas la predicción de la cotización de acciones a corto plazo

Capítulo 3

Estrategias de predicción de series de tiempo usando Deep Learning

En este capítulo, se realizará una descripción detallada de las principales técnicas de deep learning utilizadas tanto en estudios previos como en los modelos a implementar en este trabajo de fin de grado. El *deep learning* o aprendizaje profundo, como subconjunto del machine learning, se distingue por el uso de RNA (Redes Neuronales Artificiales) multicapa, las cuales ofrecen una mayor precisión en diversas tareas. Debido al éxito de los modelos de deep learning en la resolución de problemas de clasificación y regresión, la investigación continúa avanzando en la implementación de diversas arquitecturas para una amplia gama de aplicaciones, incluyendo la predicción de series temporales. Como se ha podido observar a lo largo del capítulo dedicado al estado del arte, una arquitectura de modelos en particular ha ganado popularidad a lo largo de los años debido a su efectividad en el manejo de dependencias temporales a corto plazo y su capacidad para trabajar con datos de series temporales sin procesamiento exhaustivo de características ni preprocesamiento de datos. Esta arquitectura, conocida como RNN, se ha convertido en una elección destacada para la predicción de series temporales en diversos dominios.

3.1. Redes Neuronales Recurrentes

En un nivel más detallado, las RNNs son un tipo de modelo de deep learning ampliamente utilizado en el procesamiento de datos secuenciales, lo que las convierte en una herramienta fundamental para abordar la tarea de previsión de series temporales. A diferencia de las redes neuronales Feed-forward (FFNN), donde la información se propaga en una sola dirección, las RNNs incorporan la noción de memoria y la capacidad de procesar secuencias de datos de longitud variable. El funcionamiento de las RNNs se basa en el uso de múltiples unidades recurrentes. Estas unidades comparten los mismos conjuntos de parámetros a lo largo de la secuencia de datos y procesan las entradas secuenciales en diferentes momentos de tiempo.

Cada unidad toma una entrada y produce una salida, la cual se utiliza para actualizar su estado interno. Esta actualización de estado permite a la red capturar información contextual y considerar la secuencia completa de datos en lugar de centrarse únicamente en los datos más recientes. La capacidad de las RNNs para manejar dependencias temporales a largo plazo y modelar relaciones complejas en datos secuenciales ha sido fundamental para su éxito en diversas aplicaciones, incluyendo el procesamiento del lenguaje natural, la traducción automática, la generación de texto y, por supuesto, la predicción de series temporales en el ámbito financiero.

Específicamente, el modelo RNN actualiza su memoria, también conocida como su estado oculto recurrente h_t , de la siguiente manera:

$$h_t = \sigma(W_x x_t + W_h h_{t-1} + b_t), \quad (3.1)$$

donde $x = (x_1, x_2, \dots, x_t)$ es una secuencia de longitud t , σ es una función de activación no lineal como la sigmoide o la ReLu (Unidad Lineal Rectificada), W_x y W_h son matrices de pesos, y b_t es el valor de sesgo.

Aunque las RNNs son una poderosa herramienta para el procesamiento de datos secuenciales, sufre de desafíos inherentes que pueden dificultar su entrenamiento eficiente, como el problema de la desaparición o explosión del gradiente. El gradiente en las redes neuronales es una medida de la tasa de cambio de la función de costo en relación con los parámetros del modelo de la red neuronal. Durante el entrenamiento, el gradiente se utiliza para ajustar los pesos y sesgos de la red con el objetivo de minimizar la función de costo. El problema de la desaparición del gradiente se refiere a la situación en la que el gradiente se vuelve extremadamente pequeño a medida que se propaga hacia atrás a través de capas sucesivas de la red. Esto puede llevar a que las actualizaciones de los pesos sean insignificantes, lo que dificulta que la red aprenda dependencias a largo plazo. Por otro lado, la explosión del gradiente ocurre cuando el gradiente se vuelve demasiado grande, lo que puede hacer que los pesos de la red se actualicen de manera excesiva, llevando a una convergencia ineficiente o incluso al colapso del entrenamiento.

Para superar estos desafíos, se han desarrollado variantes de las RNNs, siendo una de las más populares las Redes LSTM. Las Redes LSTM abordan el problema de la desaparición y explosión del gradiente mediante la incorporación de unidades de memoria adicionales, conocidas como celdas de memoria LSTM, que tienen la capacidad de recordar y olvidar información a lo largo del tiempo.

3.2. Long Short-Term Memory

Las redes LSTM, introducidas en 1997, son un tipo de RNN utilizado para procesar datos secuenciales y capturar dependencias a lo largo de dichos datos. A diferencia de las RNN

convencionales, que a menudo tienen dificultades para retener información de eventos distantes en el tiempo, las LSTM emplean una estructura interna especial que les permite retener información relevante durante períodos prolongados.

Como puede verse en la Figura 3.1, la estructura de una LSTM difiere de las unidades tradicionales de una red neuronal estándar y de una red neuronal recurrente, al estar compuesta de células de memoria interconectadas, diseñadas para preservar la información a largo plazo. Cada célula de memoria cuenta con tres compuertas: una compuerta de entrada, una compuerta de salida y una compuerta de olvido, [staudemeyer2019understanding](#). La compuerta de entrada controla la cantidad de información nueva que se almacena en la célula, mientras que la compuerta de salida regula la cantidad de información recuperada de la célula de memoria. La compuerta de olvido se encarga de descartar información irrelevante o no deseada. Las ecuaciones que definen estas compuertas de entrada y salida son:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f), \quad (3.2)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i), \quad (3.3)$$

$$k_t = \tanh(W_k \cdot [h_{t-1}, x_t] + b_k), \quad (3.4)$$

$$c_t = f_t x c_{t-1} + i_t x k_t, \quad (3.5)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o), \quad (3.6)$$

$$h_t = o_t * \tanh(c_t), \quad (3.7)$$

donde f_t , i_t , k_t , o_t son los valores de salida de la puerta de olvido, puerta de entrada, señal de actualización y puerta de salida, respectivamente, y sus valores de entrada son x_t en el tiempo actual t y el valor de salida h_{t-1} en el tiempo $t - 1$. W_f , W_i , W_k , W_o son matrices de pesos y b_f , b_i , b_k , b_o son vectores de sesgo correspondientes correspondiente a cada compuerta. Se considera a Sigma (σ) como una función de activación no lineal. Durante el proceso de entrenamiento, los pesos y sesgos de la red LSTM se ajustan mediante técnicas de optimización, como el descenso de gradiente estocástico (SGD), o el optimizador Adam, con el objetivo de minimizar una función de costo que mide la discrepancia entre las predicciones del modelo y los valores reales.

Las LSTM ofrecen diversas ventajas que las hacen destacar en el procesamiento de secuencias de datos, (Landi, Baraldi, Cornia, y Cucchiara, 2021). En primer lugar, su capacidad para capturar dependencias a largo plazo les permite comprender relaciones complejas en los

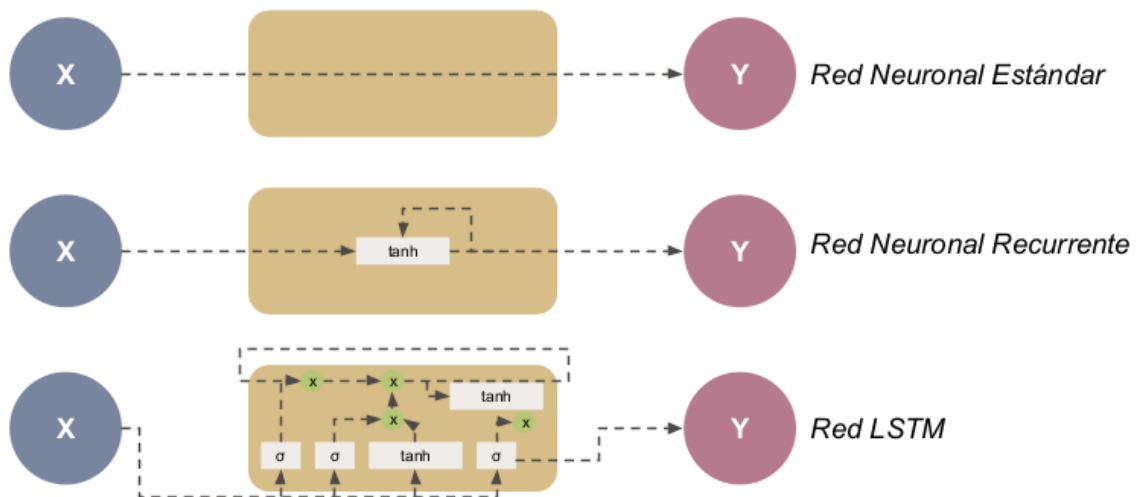


Figura 3.1: Esquema comparativo de la arquitectura de Redes Neuronales Estándar, Recurrentes y LSTM, donde X es la entrada e Y la salida correspondiente
Elaboración Propia

datos a lo largo del tiempo. Además, su estructura interna de memoria les permite retener y acceder a información relevante de entradas anteriores, facilitando la captura de contextos y patrones significativos. Sin embargo, es importante tener en cuenta que las LSTM también presentan algunas desventajas. Su mayor complejidad computacional en comparación con las redes neuronales recurrentes convencionales puede requerir más recursos y tiempo de entrenamiento.

Dada la competencia de las LSTM para manejar tanto dependencias temporales a largo como a corto plazo, el análisis del presente trabajo de fin de grado se centra en la utilización de un conjunto de modelos basados en LSTM con el objetivo de lograr una predicción más precisa del precio de las acciones a corto plazo. La elección de estas arquitecturas se basa en su capacidad demostrada para capturar las complejidades de los datos temporales y proporcionar resultados más precisos en comparación con otros enfoques.

3.3. Bidirectional Long Short-Term Memory

Una BiLSTM es una arquitectura de red neuronal que permite procesar datos secuenciales en ambas direcciones. A diferencia de una LSTM convencional, que solo tiene acceso a la información anterior en la secuencia, una BiLSTM utiliza dos redes LSTM: una que procesa la secuencia en orden ascendente y otra en orden descendente. Esta configuración bidireccional brinda a la red la capacidad de acceder tanto a la información anterior como a la posterior en la secuencia, lo que resulta beneficioso para capturar patrones más complejos y mejorar el rendimiento en tareas de predicción, graves2005frameworkise.

La arquitectura de una BiLSTM consta de dos capas LSTM, donde cada una genera una representación latente de la secuencia. Estas representaciones se concatenan en un solo vector de características, que luego se utiliza para llevar a cabo la tarea específica de la red, como la clasificación de texto o la predicción de series temporales. La capacidad de la BiLSTM para aprovechar tanto la información pasada como la futura la convierte en una opción poderosa para el análisis de datos secuenciales, permitiendo un mejor entendimiento de los patrones y una mejora en la precisión de las predicciones.

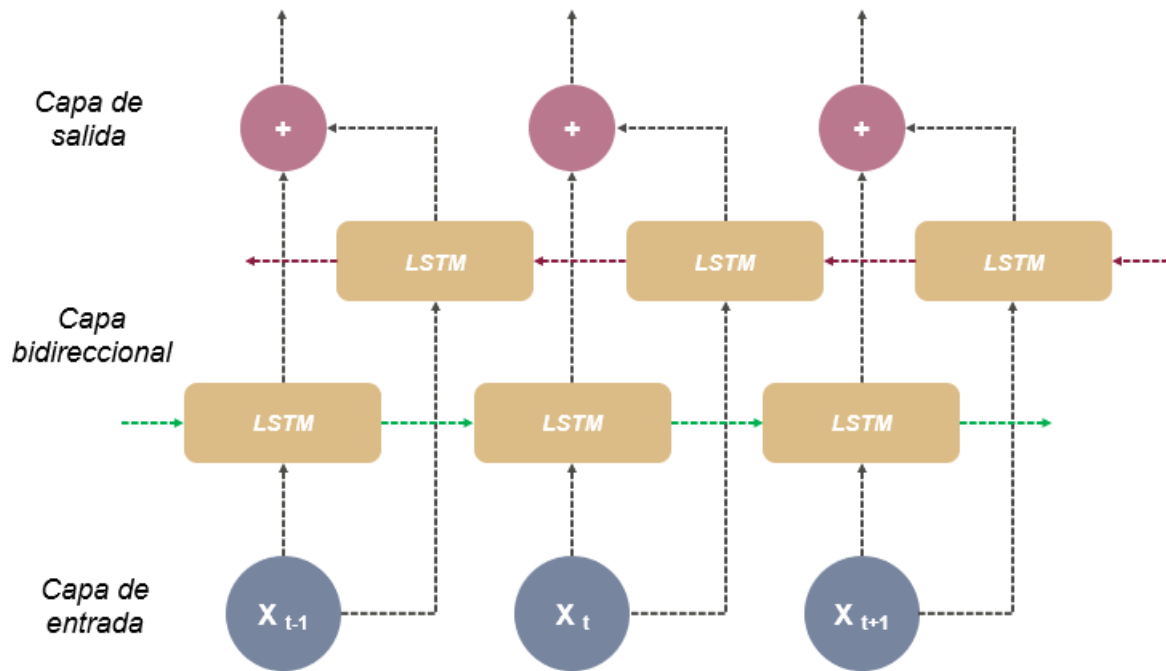


Figura 3.2: Representación gráfica de la arquitectura LSTM bilateral
Elaboración propia. Adaptada de (O.Nawajana, 2020)

La figura 3.2 ejemplifica la incorporación de la dirección para las capas anterior y posterior. La salida concatenada se define en la ecuación:

$$y_t = W \rightarrow h_t + W \leftarrow h_t + b_y, \quad (3.8)$$

donde $W \rightarrow h$ y $W \leftarrow h$ son los pesos de las redes hacia delante y hacia atrás, respectivamente, y b_y define el sesgo de la capa de salida.

La utilización de la técnica de LSTM bidireccional presenta tanto ventajas como inconvenientes. En términos de ventajas, esta arquitectura permite capturar tanto las dependencias a largo plazo como las dependencias a corto plazo en secuencias de datos, *grella2018taking*. Al procesar la secuencia en ambas direcciones, la LSTM bidireccional puede comprender y aprovechar de manera más completa el contexto de la información al considerar tanto la información pasada como la futura. Esto resulta en una capacidad mejorada para realizar predicciones más precisas y capturar patrones más complejos en los datos.

Sin embargo, es importante tener en cuenta algunas de las desventajas asociadas con el uso de una LSTM bidireccional. Una de las principales desventajas es su mayor complejidad computacional en comparación con las arquitecturas unidireccionales. Debido a que la red procesa la secuencia en dos direcciones simultáneamente, esto requiere un mayor número de cálculos y, en consecuencia, puede demandar más recursos y tiempo de entrenamiento.

Además, la combinación de información tanto pasada como futura puede dificultar la interpretación de los resultados y la comprensión de los procesos internos de la red, lo que puede suponer un desafío adicional para los analistas al tratar de entender y explicar el funcionamiento de la red. A pesar de estos inconvenientes, la LSTM bidireccional sigue siendo una herramienta valiosa al analizar series temporales debido a su capacidad mejorada para modelar contextos complejos y realizar predicciones más precisas.

3.4. Stacked Long Short-Term Memory

El Stacked LSTM es una variante de la red neuronal recurrente LSTM que presenta múltiples capas apiladas una encima de la otra (Ver Figura 3.3). Cada capa en la pila produce una representación latente de la secuencia que se utiliza como entrada para la siguiente capa. La principal idea detrás del Stacked LSTM radica en la capacidad de cada capa para aprender representaciones más complejas y abstractas a medida que se agregan más capas. Esta estructura proporciona una serie de ventajas destacadas, ya que:

1. **Mayor capacidad de representación:** Al apilar varias capas LSTM, el Stacked LSTM tiene la capacidad de aprender representaciones más profundas y complejas de los datos secuenciales. Al capturar información en diferentes niveles de abstracción, la red puede identificar patrones más sutiles y sofisticados en los datos. Esto es especialmente valioso en tareas de procesamiento de lenguaje natural (NLP), reconocimiento de voz y análisis de series temporales, donde existen dependencias y relaciones complejas.
2. **Mejora de la precisión:** En comparación con una LSTM individual, los modelos Stacked LSTM pueden lograr una mejora significativa en la precisión. Al tener múltiples capas, la red tiene la capacidad de aprender representaciones más refinadas y detalladas de los datos secuenciales, lo que se traduce en una mayor capacidad para realizar predicciones precisas.
3. **Reducción del sobreajuste (*overfitting*):** Como ocurre con cualquier red neuronal profunda, los modelos Stacked LSTM pueden estar sujetos al sobreajuste, donde la red se ajusta demasiado a los datos de entrenamiento y no generaliza bien a nuevos datos. Sin embargo, al agregar capas adicionales, se permite que la red aprenda patrones más complejos y se reduzca el riesgo de sobreajuste. Esto se debe a que cada capa en la pila proporciona una oportunidad adicional para extraer características y representaciones

útiles de los datos, evitando así que la red se sobreajuste a patrones específicos del conjunto de entrenamiento.

Por otro lado, la arquitectura de Stacked LSTM presenta algunas desventajas importantes. En primer lugar, su mayor complejidad computacional y tiempo de entrenamiento pueden limitar su aplicabilidad en entornos con recursos limitados. En segundo lugar, la adición de capas adicionales aumenta la complejidad interpretativa de la red, dificultando la comprensión de los procesos internos y la explicación de los resultados. Por último, la presencia de más hiperparámetros en la red requiere un ajuste y optimización cuidadosos, lo que puede aumentar la complejidad en la búsqueda de los valores óptimos.

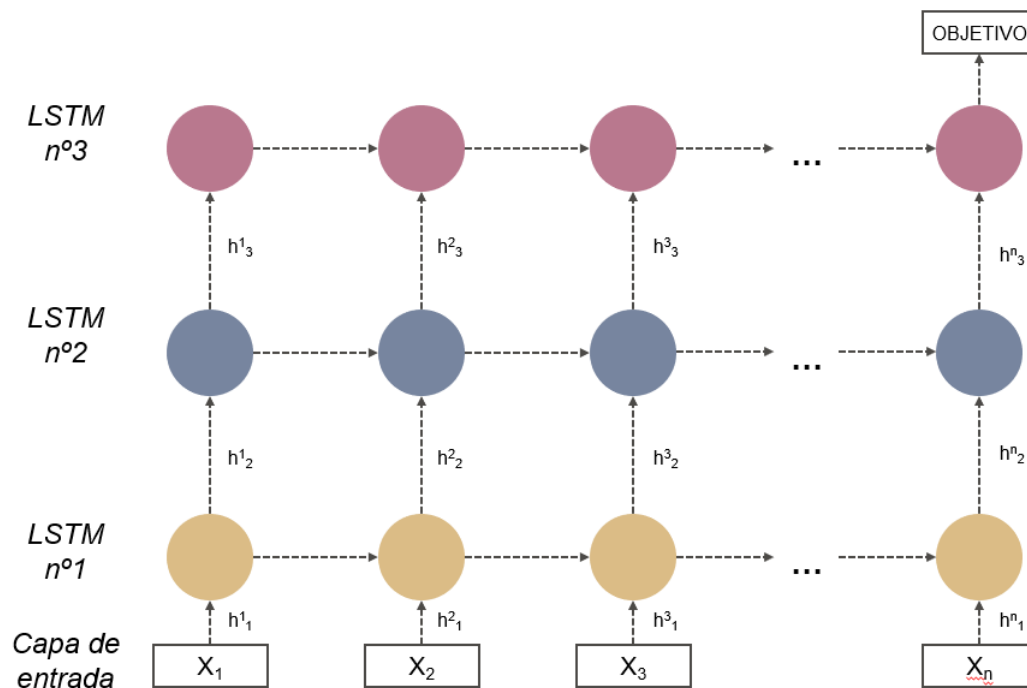


Figura 3.3: Estructura jerárquica de un Stacked LSTM
Elaboración Propia

3.5. Convolutional Neural Network

Las CNN son arquitecturas de redes neuronales profundas ampliamente utilizadas en tareas de procesamiento de imágenes y videos. A diferencia de las redes neuronales tradicionales, las CNN aprovechan operaciones matemáticas como la convolución y la agrupación para procesar datos con relaciones espaciales, como imágenes. Estas redes cuentan con capas convolucionales, encargadas de aplicar filtros para extraer características relevantes de los datos.

Estas capas se combinan con capas de agrupación, que reducen el tamaño de la imagen para una mayor eficiencia de procesamiento. Como puede verse en la Figura 3.4, el diseño

de las CNN sigue este patrón hasta que la imagen se reduce lo suficiente para ser procesada por capas de neuronas completamente conectadas, que realizan la clasificación final.

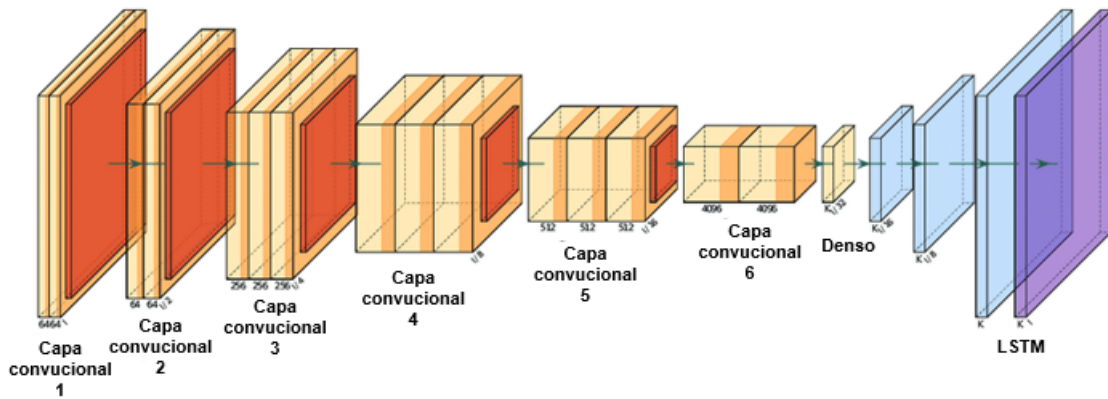


Figura 3.4: Ejemplo de una CNN con 6 capas
Elaboración Propia. Adaptada de (Delteil, 2023)

Las CNN ofrecen numerosas ventajas en el procesamiento de imágenes, salehi2023study. Una de las principales es su capacidad para aprender características espaciales automáticamente, sin la necesidad de que el usuario especifique manualmente dichas características, como bordes o texturas. Durante el entrenamiento, la CNN optimiza sus parámetros para identificar y reconocer estas características de forma eficiente. Sin embargo, las CNN presentan algunas limitaciones. Una de ellas es su dificultad para capturar relaciones a largo plazo en secuencias de datos, lo cual las hace menos adecuadas para tareas basadas en secuencias, como el procesamiento del lenguaje natural o la predicción de series de tiempo. Además, las CNN requieren conjuntos de datos de entrenamiento lo suficientemente grandes y representativos para aprender características generalizables, lo cual puede ser un desafío en dominios con datos escasos o desequilibrados. Estas desventajas señalan la necesidad de considerar otras arquitecturas, como las redes recurrentes, para tareas específicas que involucren dependencias temporales complejas.

3.6. Convolutional Neural Network Long Short-Term Memory

Un CNN LSTM es una arquitectura de red neuronal que combina los beneficios de las CNN y las LSTM para procesar datos secuenciales de manera más efectiva. Al incorporar las estructuras convolucionales de las CNN tanto en las transiciones de entrada a estado como de estado a estado, el CNN LSTM es capaz de capturar tanto las características espaciales

como las dependencias temporales en los datos secuenciales. La arquitectura de un CNN LSTM consta de dos partes principales: la parte CNN, encargada de extraer características significativas de los datos de entrada mediante la aplicación de filtros convolucionales, y la parte LSTM, que procesa las características extraídas y modela las dependencias a largo plazo en la secuencia lu2020cnn. Esta combinación de CNN y LSTM se representa de forma esquemática en la Figura 3.5, mostrando cómo las capas convolucionales y recurrentes se integran en una sola arquitectura. Al unir las capacidades de extracción de características espaciales de las CNN con la capacidad de modelar dependencias temporales de las LSTM, el CNN LSTM se ha utilizado con éxito en una variedad de tareas, como reconocimiento de acciones en videos, análisis de series temporales y generación de descripciones de imágenes.

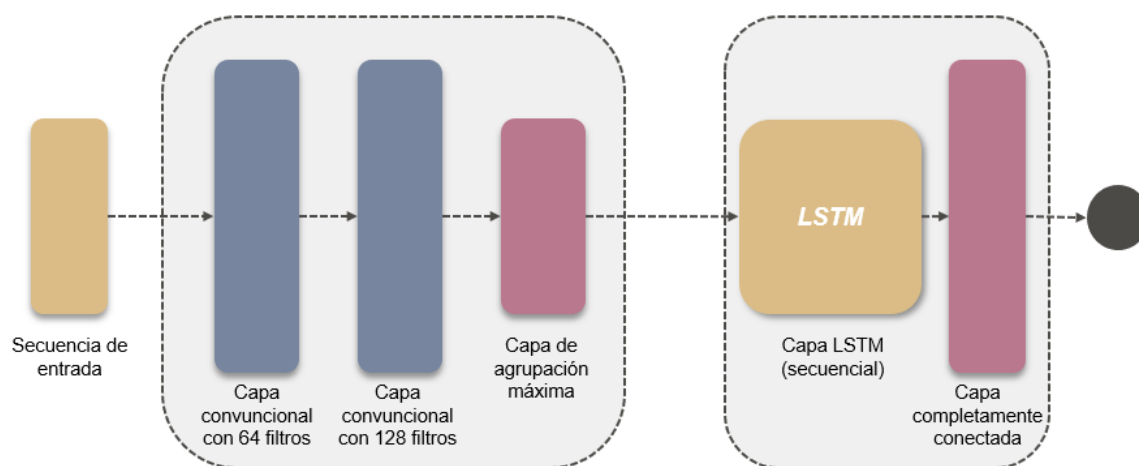


Figura 3.5: Arquitectura CNN-LSTM
Elaboración Propia

La arquitectura de la CNN LSTM presenta diversas ventajas en comparación con enfoques individuales como las CNN o las LSTM li2020hybrid. Una de las principales ventajas radica en la capacidad de combinar la extracción de características espaciales de las CNN con la captura de dependencias a largo plazo de las LSTM. Esta combinación permite a la CNN LSTM aprender patrones temporales y características espaciales simultáneamente, lo que resulta especialmente útil en tareas como la generación de subtítulos automáticos y el análisis de videos en tiempo real. Al incorporar tanto información espacial como temporal en su proceso de aprendizaje, la CNN LSTM es capaz de realizar un análisis más completo y sofisticado de los datos secuenciales.

No obstante, una de las desventajas asociadas a la CNN LSTM es el ajuste adecuado de los hiperparámetros y la gestión del equilibrio entre las capas convolucionales y las capas LSTM. Al combinar estas dos técnicas, se añade una mayor complejidad al proceso de configuración

y ajuste de la red. Es crucial encontrar el equilibrio adecuado entre las capas convolucionales y las capas LSTM para evitar un sesgo excesivo hacia las características espaciales o las dependencias temporales. Además, la selección óptima de hiperparámetros, como el tamaño de los filtros convolucionales y el número de capas LSTM, puede resultar más desafiante debido a la combinación de ambas técnicas

3.7. Métricas de evaluación

Una vez que se han desarrollado los diferentes modelos, es fundamental evaluar su eficacia y precisión mediante el uso de diversas métricas de evaluación. Algunas de las métricas más comúnmente utilizadas son las siguientes:

1. **RMSE:** esta métrica se calcula tomando la raíz cuadrada del promedio de los errores al cuadrado entre las predicciones del modelo y los valores reales. Su fórmula se expresa de la siguiente manera:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_{\text{pred}}^{(i)} - y_{\text{actual}}^{(i)})^2} \quad (3.9)$$

2. **MAE:** esta métrica se calcula como la diferencia promedio entre los valores reales y las predicciones del modelo, ignorando la dirección de las desviaciones. Su fórmula se expresa de la siguiente manera:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_{\text{actual}}^{(i)} - y_{\text{pred}}^{(i)}| \quad (3.10)$$

El MAE se expresa en la misma unidad que la variable objetivo, lo que facilita la interpretación. Un valor de MAE más bajo indica que el modelo tiene una mayor precisión, ya que las predicciones están más cercanas a los valores reales. Por otro lado, un MAE más alto indica una menor precisión del modelo.

3. **R²:** este coeficiente representa la proporción de la varianza de la variable dependiente que es explicada por el modelo. Su valor puede variar entre 0 y 1. Un valor de 1 indica que el modelo es capaz de explicar toda la variabilidad de los datos, mientras que un valor de 0 indica que el modelo no puede explicar ninguna variabilidad y simplemente se ajusta a la media de los datos. Se calcula mediante la siguiente fórmula:

$$R^2 = 1 - \frac{SSE}{SST} \quad (3.11)$$

donde SSE (Sum of Squares of Errors) es la suma de los cuadrados de las diferencias

entre las predicciones del modelo y los valores reales, y SST (Total Sum of Squares) es la suma de los cuadrados de las diferencias entre los valores reales y la media de los datos.

Capítulo 4

Predicción a corto plazo de acciones usando Deep Learning

4.1. Esquema Metodológico

La parte experimental de este proyecto se centra en la predicción de la cotización de las acciones de una empresa utilizando técnicas de deep learning, descritas en detalle en el Capítulo 3. Los códigos usados para el análisis y desarrollo de los modelos han sido incluidos en el siguiente repositorio de github, (Gallego, 2023). Con base en esta información, los pasos seguidos para la realización de este trabajo son:

1. **Adquisición y pre-procesamiento de datos :** En esta etapa, se recopilan los datos históricos de la cotización de acciones de la empresa de interés, en este caso ACCIONA. Estos datos pueden obtenerse de diversas fuentes, como bases de datos financieras o servicios de proveedores especializados. En este trabajo, se ha usado Yahoo Finance como fuente de datos (Acciona Yahoo, 2023). Una vez recopilados, se cargan en el entorno de trabajo para su posterior procesamiento. Asimismo, es fundamental realizar una limpieza exhaustiva de los datos antes de utilizarlos en el modelo. Esto implica manejar y solucionar problemas comunes como datos faltantes, valores atípicos y duplicados. Además, es posible que se requiera ajustar la escala de los datos o aplicar técnicas de normalización para garantizar un buen rendimiento del modelo.
2. **Análisis Descriptivo de datos:** Antes de implementar los modelos de redes neuronales, es importante comprender en profundidad las características de la serie temporal de cotizaciones de acciones. Esto implica analizar la distribución de los datos, identificar patrones estacionales o tendencias, y explorar posibles correlaciones con otras variables relevantes. El análisis exploratorio proporciona información clave para la elección y diseño adecuado de los modelos.

3. **Implementación de modelos de redes neuronales:** Teniendo en cuenta la revisión del estado del arte en el campo detallado en el capítulo 2, se implementará una variedad de modelos de redes neuronales, adaptados a las características específicas de la serie temporal analizada. Se llevará a cabo la implementación de las cuatro técnicas principales mencionadas en el capítulo 3: LSTM, Stacked LSTM, BiLSTM y CNN-LSTM. Además, se realizarán pruebas con modificaciones en los hiperparámetros, como la longitud de la ventana temporal al crear los conjuntos de datos para el proceso de entrenamiento y prueba, así como el número de neuronas, para cada una de las técnicas desarrolladas.
4. **Evaluación y comparación de modelos:** Tras la implementación de 90 variaciones de las 4 técnicas mencionadas previamente, se procederá a evaluar la precisión de cada una de ellas utilizando métricas de evaluación clave como RMSE, MAE y R2. Mediante esta comparativa, se identificará la combinación más efectiva para cada técnica. A continuación, se llevará a cabo una evaluación adicional para determinar la mejor combinación global, considerando factores importantes para el modelo, como la longitud de la ventana temporal y el número de neuronas. La elección final se basará en la combinación que demuestre un rendimiento superior en términos de precisión y capacidad de generalización en los conjuntos de datos de entrenamiento y prueba.

En este contexto, en la figura 4.1, se esquematiza la metodología implementada en este trabajo, proporcionando una visión general de las etapas y pasos a seguir en el proceso experimental.

4.2. Adquisición de datos y pre-procesamiento

Para desarrollar herramientas efectivas que estimen los valores futuros de las cotizaciones de acciones, es de vital importancia contar con medidas de alta calidad de los datos de precios históricos. La calidad y la cantidad de los datos desempeñan un papel fundamental en los resultados de las estimaciones y pueden dar lugar a suposiciones incorrectas si no se adquieren de manera adecuada. Por lo tanto, la selección de un conjunto de datos completo y rico se convierte en una tarea determinante para garantizar la precisión y la fiabilidad de los modelos predictivos. El primer paso para asegurar la calidad de los datos es garantizar que se disponga de una amplia gama de observaciones históricas. Cuanto mayor sea la cantidad de datos disponibles, mayor será la capacidad de los modelos para capturar las variaciones y los patrones inherentes a la serie temporal de precios. Además, es esencial que los datos estén completos, es decir, que no haya lagunas o valores faltantes, ya que esto puede distorsionar las estimaciones y afectar la precisión de los modelos.

En el marco de este estudio, se ha seleccionado una empresa del IBEX 35, el principal índice bursátil español, para utilizar sus datos en el modelo de predicción. En particular, se



Figura 4.1: Enfoque metodológico utilizado para llevar a cabo la predicción del valor de cierre de una acción
Elaboración Propia

analizarán los precios históricos de las acciones de Acciona, una destacada empresa española especializada en la promoción y gestión de infraestructuras y energías renovables. Acciona ha sido elegida debido a su relevancia tanto a nivel nacional como internacional. Acciona ocupa una posición destacada en el IBEX, con una capitalización bursátil de 8.52 mil millones de euros, una facturación de más de 11 mil millones de euros en 2022 y un beneficio neto de 441 millones de euros, (Fariza, 2023). Además de su importancia en el ámbito nacional, la empresa tiene presencia en sesenta y cinco países de los cinco continentes, lo que demuestra su alcance global en el sector, (Acciona, 2023).

Los datos históricos de los precios de las acciones de Acciona se obtuvieron de la página web de Yahoo Finance, una fuente reconocida y ampliamente utilizada para obtener información financiera, (Acciona Yahoo, 2023). Con el objetivo de garantizar una predicción precisa, se decidió recopilar al menos cinco años de información, lo que permitirá capturar las tendencias y patrones a largo plazo en los precios de las acciones. Los datos abarcan un período que va desde el 26 de junio de 2015 hasta el 19 de junio de 2023, con una frecuencia diaria. Esta elección temporal proporciona una amplia cobertura y una visión significativa de la evolución de los precios de las acciones de Acciona a lo largo del tiempo, lo que resulta fundamental para el análisis y la predicción. Al realizar la descarga de los datos, se obtiene una amplia gama de información relevante relacionada con los precios de cotización de la compañía. La Tabla 4.1 proporciona una lista detallada de las variables y su significado

correspondiente, lo que facilita la comprensión de los datos y su interpretación adecuada.

Variable	Definición
Date	Fecha de cada valor
Open	Precio de apertura de la acción en cada una de las fechas
Close	Precio de cierre de la acción en cada una de las fechas ajustado por splits
Adj. close	Precio de cierre ajustado por divisiones, dividendos y plusvalías
High	Valor máximo del precio de la acción en cada jornada
Low	Valor Mínimo del precio de la acción en cada jornada
Volume	Cantidad de transferencias de compraventa durante una jornada
Year	Año de la fecha
Month	Mes del año de la fecha
Day	Día del mes
Weekday	Del 0 al 4, de lunes a viernes, el día de la semana
Cumulative return	Retorno acumulado desde el inicio del periodo temporal analizado

Tabla 4.1: Definición de las variables del dataset

4.3. Análisis Descriptivo

Antes de iniciar el análisis, es importante obtener una primera perspectiva de los datos que se van a analizar. En la figura 4.2, se presenta un gráfico de velas que muestra los datos históricos del precio de las acciones de la compañía Acciona. Los gráficos de velas son una herramienta visual ampliamente utilizada en el análisis técnico de los mercados financieros, especialmente en el trading de acciones, divisas y otros instrumentos financieros. Cada vela en el gráfico proporciona información detallada sobre los precios de apertura, cierre, máximo y mínimo de un activo durante un período de tiempo específico.

Específicamente, la estructura de una vela consiste en un cuerpo y dos sombras. El cuerpo representa la diferencia entre el precio de apertura y el precio de cierre del activo en ese período, con una frecuencia diaria. Si el precio de cierre es mayor que el precio de apertura, el cuerpo de la vela es de color verde, lo que indica un movimiento alcista. Por el contrario, si el precio de cierre es menor que el precio de apertura, el cuerpo de la vela es de color rojo, lo que indica un movimiento bajista. Las sombras, también conocidas como mechas o colas, representan la diferencia entre los precios máximos y mínimos alcanzados durante ese período. Como se observa en la figura 4.2, los datos históricos del precio de cierre de las acciones de Acciona, muestran una tendencia alcista a nivel general, con especial repunte en las fechas de finales de 2020, principios del 2021 y finales del primer trimestre del 2022.

A continuación, se realizará un análisis descriptivo exhaustivo de los datos con el fin de comprender la dinámica asociada al precio de cierre de las acciones. Como primer paso, se procederá a transformar la variable “Date” en un objeto del DataFrame para que sea reconocida como una fecha y se pueda manipular adecuadamente durante el proceso de predicción.

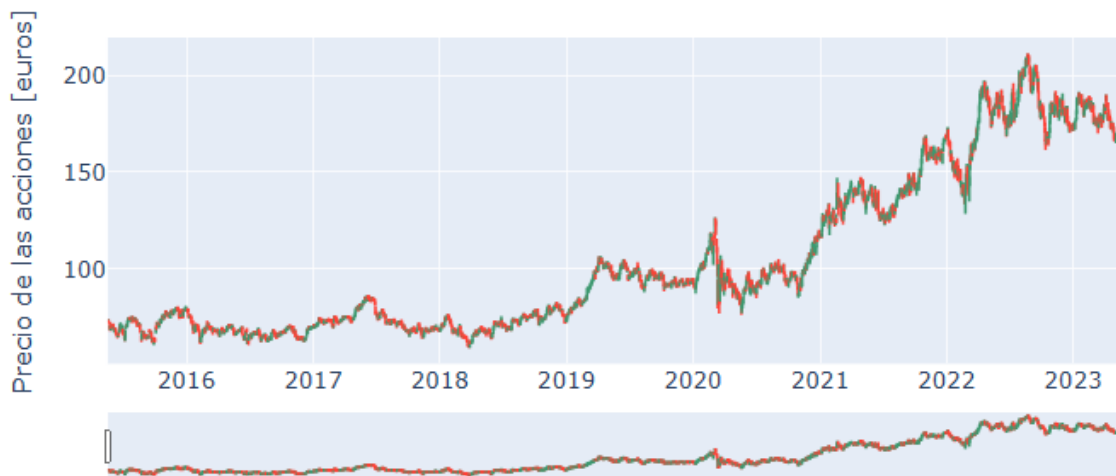


Figura 4.2: Gráfico de Vela de los precios de apertura y cierre de las acciones de Acciona
Elaboración Propia

El conjunto de datos consta de un total de 2050 observaciones diarias, abarcando un período significativo de tiempo. Antes de continuar con el análisis descriptivo, es esencial identificar y tratar los posibles valores nulos en los datos. Sin embargo, al examinar detalladamente el conjunto de datos, se ha verificado que no existen valores faltantes o nulos en ninguna de las observaciones. Esta ausencia de datos faltantes es un aspecto positivo, ya que garantiza la integridad de los registros y evita la necesidad de realizar imputaciones o estimaciones.

Para obtener una comprensión más profunda de los datos y evaluar su calidad, se calcularán los principales estadísticos de las variables seleccionadas. Estas estadísticas incluyen el recuento de observaciones, la media, el valor máximo, el valor mínimo, la desviación estándar y los cuartiles. Este análisis proporcionará información clave que nos permitirá comprender la distribución de los datos, detectar posibles valores atípicos y evaluar la escala de las variables. En este sentido, la Tabla 4.2 incluye las estadísticas descriptivas del conjunto de datos analizado.

Estadísticos	Open	High	Low	Close	Volume	Retorno
Recuento de datos	2050	2050	2050	2050	2050	2050
Media	104.36	105.53	103.06	104.35	1.50e+05	0.00
Desviación Estándar	41.24	41.81	40.65	41.28	1.37e+05	0.02
Mínimo	59.66	60.50	59.44	59.82	0.00	-0.17
25 %	71.09	71.73	70.37	70.96	8.31e+04	-0.01
50 %	90.65	91.79	89.25	90.77	1.18e+05	0.00
75 %	135.19	137.35	133.30	135.30	1.77e+05	0.01
Máximo	210.60	211.00	206.80	210.20	2.74e+06	0.15

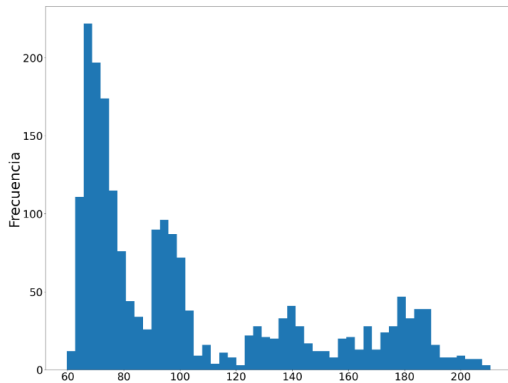
Tabla 4.2: Estadísticas de las variables del conjunto de datos analizado

Como se observa en la Tabla 4.2, y cómo es lógico ya que no existen valores nulos como se ha mostrado anteriormente, todas las variables cuentan con 2050 observaciones. Las variables de precio de apertura, cierre, máximo y mínimo tienen estadísticos muy similares ya que el precio de una acción no debería variar de forma significativa diariamente. Además, no se observan valores significativamente atípicos, tanto máximos como mínimos tienen una dispersión común con respecto a la media. Además, todos estos valores son positivos ya que un precio puede reducirse, pero nunca podrá ser menor que cero. Por su parte, la variable 'Return' si tiene valores tanto positivos como negativos en función de si el precio de cierre es mayor o menor al de apertura.

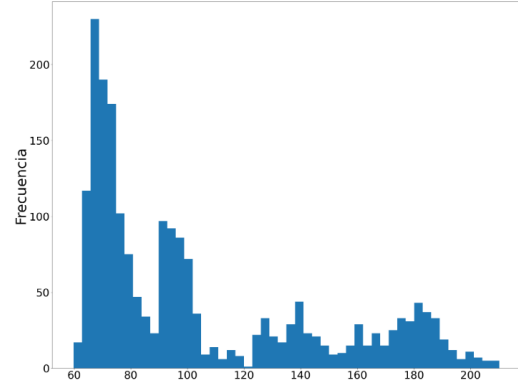
Asimismo, para analizar la distribución de valores de cada variable en mayor detalle, se han obtenido sus histogramas. Estos gráficos proporcionan una visualización clara de la frecuencia de los valores en cada rango. En la figura 4.3, se muestran los histogramas correspondientes a los diferentes precios históricos de las acciones de Acciona y al retorno. En el caso del precio histórico de cierre las acciones, se observa que la mayor concentración de valores se encuentra en el rango de 60 a 80 euros por acción. La distribución de los precios históricos presenta una asimetría hacia la derecha, lo que significa que existen precios de cierre más altos y menos frecuentes en comparación con los precios más bajos y más comunes. En general, indica que hay una cola larga hacia la derecha, lo que implica que hay eventos o condiciones que ocasionalmente generan valores muy altos de precios de cierre. Por otro lado, el histograma del retorno muestra una distribución significativamente normal. Esto implica que la mayoría de los retornos se concentran alrededor de la media y disminuyen a medida que nos alejamos de ella. Los retornos más altos y más bajos son menos probables y se vuelven más raros a medida que nos alejamos de la media. Es importante destacar que estos resultados son consistentes con la naturaleza volátil de los mercados financieros. Los cambios rápidos y los eventos imprevistos pueden afectar los precios de las acciones y provocar desviaciones de una distribución normal. Sin embargo, en el caso de los datos analizados, se observa una distribución aproximadamente normal en el retorno, lo que indica una estabilidad relativa en el comportamiento de los precios en el período considerado.

Antes de implementar el modelo utilizando diferentes técnicas de deep learning, es necesario realizar la normalización de las variables. La normalización implica transformar los valores de una variable de modo que su distribución se ajuste a una escala específica, generalmente entre 0 y 1. Esta etapa es fundamental en estos modelos de redes neuronales porque la normalización facilita la convergencia del modelo durante el entrenamiento. Al tener todas las variables en una escala similar, los algoritmos de optimización pueden encontrar de manera más eficiente los mínimos locales y globales en la función de pérdida, lo que mejora la velocidad y estabilidad del proceso de entrenamiento.

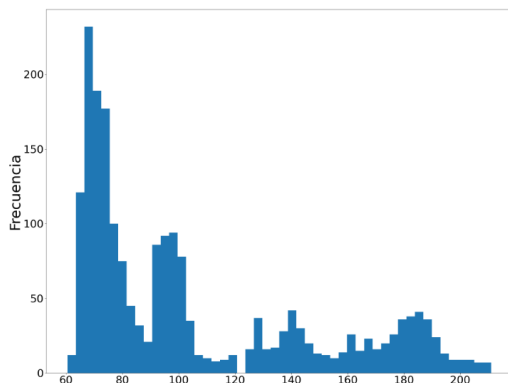
Debido al enfoque de este trabajo en la predicción univariante y en concordancia con los resultados obtenidos en el Capítulo 2, se ha decidido utilizar el precio de cierre como la variable principal en los modelos de predicción. Esta elección se basa en varios factores



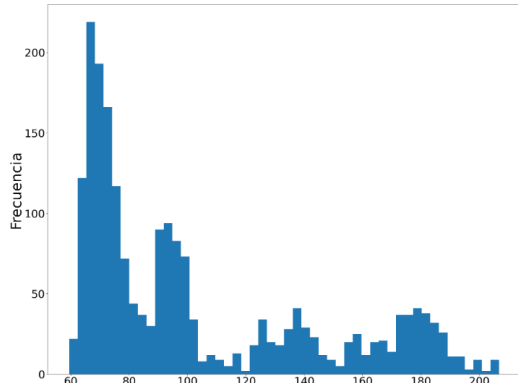
(a) Histograma del precio de apertura



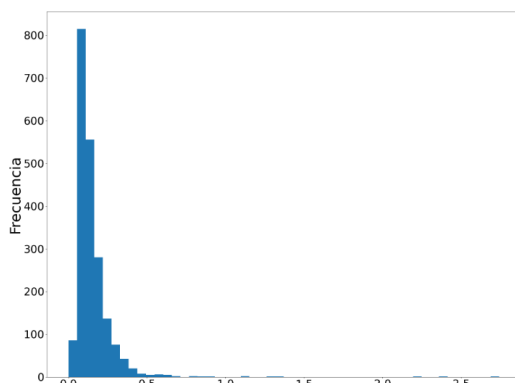
(b) Histograma del precio de cierre



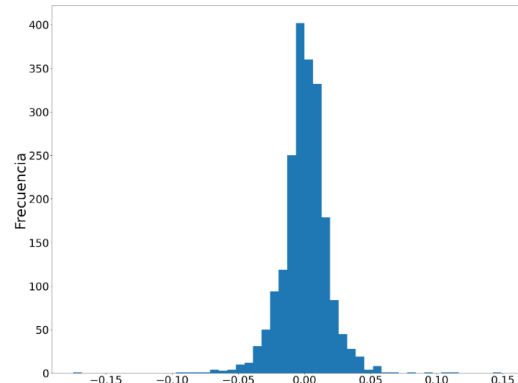
(c) Histograma del precio de máximo



(d) Histograma del precio de mínimo



(e) Histograma del volumen



(f) Histograma del retorno

Figura 4.3: Distribución de las variables obtenidas en el data-set

que respaldan la relevancia y la utilidad de esta variable para el análisis y la estimación de los precios de las acciones. En primer lugar, el precio de cierre representa el valor final al que se ha negociado una acción en un día determinado. Al utilizar esta variable, se tiene en cuenta la información más reciente disponible en el mercado y se refleja la evaluación global de los inversores al final de la jornada. Además, el precio de cierre es ampliamente utilizado en el análisis técnico y fundamental en el ámbito financiero. Los analistas y traders suelen utilizar este valor como referencia para identificar patrones, tendencias y niveles de soporte y resistencia. Así, la serie de tiempo asociada a los precios de cierre entre los años 2015 y 2023, que se va a analizar en los modelos de predicción se observa en la figura 4.4.

Acciona



Figura 4.4: Precio de cierre a lo largo del tiempo
Elaboración Propia

En la figura 4.4, se puede apreciar la tendencia positiva de la cotización de las acciones de Acciona a lo largo del período analizado. Esta tendencia se ve reflejada en el aumento sostenido del precio de las acciones, alcanzando su punto máximo en agosto de 2022. Sin embargo, también se observan momentos significativos de fluctuación en el precio de las acciones. Uno de estos momentos se produce en el año 2020, cuando la crisis sanitaria y financiera del Covid-19 tuvo un impacto negativo en los mercados financieros a nivel mundial. Durante este período, se registra una caída en el precio de las acciones de Acciona. Sin embargo, es importante destacar que esta disminución se ve compensada por una posterior revalorización de las acciones. Un evento clave que contribuye a esta revalorización se produjo en julio de 2021, cuando Acciona Energía, la rama de energías renovables de Acciona, salió a bolsa.

Esta salida a bolsa generó un aumento en el interés y la confianza de los inversores, lo que se refleja en el incremento del precio de las acciones.

A continuación, como parte del análisis del comportamiento histórico del precio de cierre, se procede a generar gráficos de boxplot con una frecuencia mensual y anual, con el objetivo de examinar las diferencias en la dispersión de los datos. Estas representaciones visuales, mostradas en las figuras 4.5 y 4.6, nos permiten identificar patrones y tendencias en el comportamiento de los precios para diferentes frecuencias.

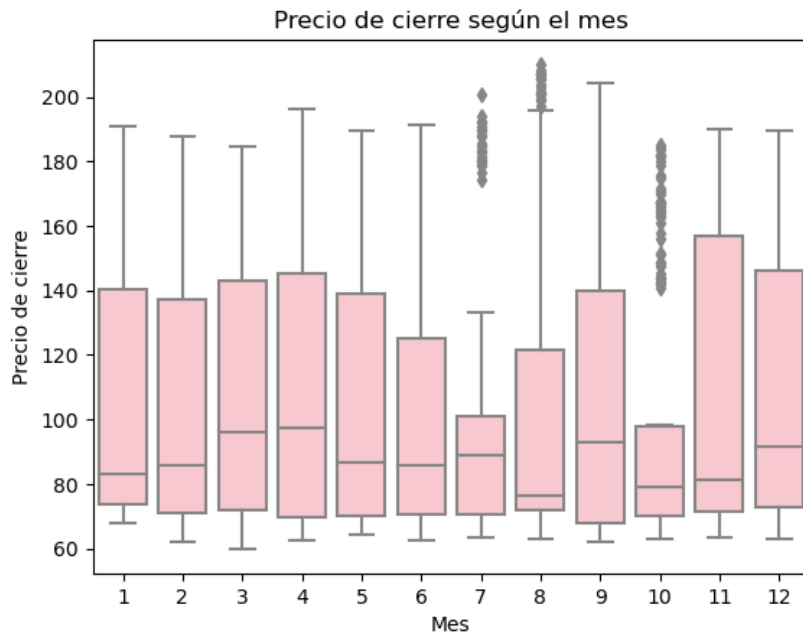


Figura 4.5: Precio de cierre a lo largo del tiempo
Elaboración Propia

Al analizar el gráfico de boxplot mensual (Ver Figura 4.5), observamos que el mes de abril presenta el mayor precio de cierre mediano en comparación con los otros meses del año. Esto sugiere que históricamente ha habido una mayor demanda o un mayor interés por las acciones de Acciona durante ese mes en particular. Por otro lado, sin tener en cuenta los datos atípicos, se identifica que el mes de octubre presenta una menor dispersión de precios en comparación con los demás meses, lo que indica una mayor estabilidad en el valor de cierre de las acciones durante ese periodo.

Al analizar la frecuencia anual en el comportamiento del precio de cierre (ver Figura 4.6), se destaca el año 2022 como un periodo con una alta dispersión de precios y la máxima cotización. Esto se debe a dos causas principales que influyeron en este aumento significativo: la revaloración de las acciones de Acciona tras la salida a bolsa de Acciona Energía y la recuperación económica posterior a la crisis del Covid-19. En primer lugar, la salida a bolsa de Acciona Energía en julio de 2021 generó un gran interés y expectativa en el mercado. Esto llevó a una mayor demanda de las acciones de Acciona, lo que se tradujo en un incremento en

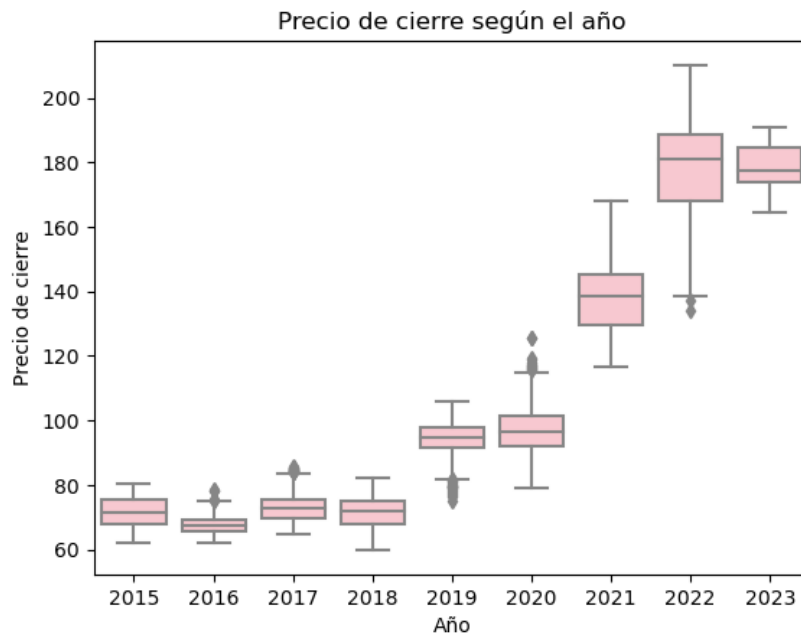


Figura 4.6: Precio de cierre a lo largo del tiempo
Elaboración Propia

el precio de cierre promedio durante el año 2022. Los inversionistas y analistas reconocieron el valor de la rama de energías renovables de Acciona, lo que impulsó el valor de las acciones y contribuyó a su revaloración, (Fariza, 2023). En segundo lugar, la recuperación económica global después de la crisis del Covid-19 tuvo un impacto positivo en los mercados financieros. Los bancos centrales y los gobiernos implementaron medidas económicas para estimular la inversión y reducir el ahorro. Estas políticas tuvieron como resultado un aumento en la demanda de acciones y una mayor participación de los inversores en los mercados bursátiles. Como resultado, los precios de las acciones, incluyendo los de Acciona, experimentaron un incremento durante el año 2022.

Después de analizar y comprender el comportamiento de los datos históricos de las acciones de Acciona, se procede a la implementación del modelo de predicción del precio de cierre utilizando diversas técnicas de Deep Learning, las cuales han sido previamente explicadas en el Capítulo 3.

4.4. Implementación de modelos

En esta sección se presenta una explicación detallada de la implementación de modelos de predicción utilizando RNN, específicamente basados en unidades LSTM. A lo largo de los Capítulos 2 y 3, se ha destacado el papel fundamental de las RNNs, en el ámbito del aprendizaje profundo, para la predicción de series temporales. Así, en primer lugar, se han cargado los datos de los precios históricos de Acciona desde junio de 2015 hasta la fecha

de descarga (junio de 2023). Estos datos se han cargado en formato CSV y se ha realizado un proceso de preprocesamiento. Durante esta etapa, se ha verificado que no hayan valores faltantes en los datos y se han normalizado para su uso en los modelos de redes neuronales. Es importante recordar que los modelos de Deep Learning implementados se centran en un análisis univariante para predecir a corto plazo (un paso adelante) el precio de cierre de las acciones de Acciona.

Para construir los conjuntos de entrenamiento y prueba, se emplea una técnica de ventana deslizante que recorre la serie temporal completa y divide los datos en observaciones. Cada ventana se considera como una secuencia de entrada (*dataX*), mientras que el valor siguiente a la ventana se toma como la salida correspondiente (*dataY*). El tamaño de esta ventana es uno de los parámetros clave en el modelo y se ajusta de acuerdo con las necesidades específicas. En nuestro caso, para una serie temporal diaria, se han explorado tamaños de ventana de 30, 60 y 90 días. Esto implica que el modelo utilizará las 30, 60 o 90 observaciones más recientes para predecir el siguiente valor. Una vez que la serie temporal se ha dividido en ventanas individuales, se procede a separarlas en un conjunto de entrenamiento y un conjunto de prueba. En este proceso, se asigna una proporción mayor de datos al conjunto de entrenamiento en comparación con el conjunto de prueba. Esta decisión se toma para garantizar que el modelo tenga suficientes datos para aprender y ajustarse durante la fase de entrenamiento, mientras se mantiene un conjunto de prueba lo suficientemente representativo para evaluar la capacidad de generalización del modelo.

Inicialmente, todas las técnicas se desarrollaron utilizando una partición del 80 % para el conjunto de entrenamiento (*training set*) y un 20 % para el conjunto de pruebas (*test set*). Sin embargo, esta metodología arrojó resultados altamente precisos en el conjunto de entrenamiento, pero poco satisfactorios en el conjunto de pruebas. Cuando los resultados en el conjunto de entrenamiento son excelentes pero no se reflejan en el conjunto de pruebas, generalmente indica que el modelo se ha ajustado demasiado o ha ocurrido un fenómeno conocido como sobreajuste (*overfitting*). El sobreajuste ocurre cuando el modelo se adapta demasiado a los detalles específicos del conjunto de entrenamiento y no logra generalizar correctamente para nuevos datos. Esto puede suceder cuando el conjunto de entrenamiento es relativamente pequeño en comparación con la complejidad del modelo, o cuando el modelo tiene una capacidad excesiva para capturar información irrelevante en el conjunto de entrenamiento.

Es fundamental abordar el problema del sobreajuste para obtener un modelo más robusto y preciso. Algunas estrategias comunes para mitigar el sobreajuste incluyen aumentar el tamaño del conjunto de entrenamiento, utilizar técnicas de regularización como la regularización L1 o L2, y aplicar estrategias como *early stopping*. En el caso de este modelo en particular, se tomó la decisión de ampliar la muestra de pruebas, cambiando la partición a un 60 % para el conjunto de entrenamiento y un 40 % para el conjunto de pruebas. Además, se implementó la estrategia de *early stopping*, comúnmente utilizada en modelos de deep

learning para prevenir el sobreajuste y mejorar el rendimiento del modelo (X.-x. Wu y Liu, 2009). Esta estrategia consiste en detener el entrenamiento del modelo antes de que alcance su convergencia completa, basándose en el seguimiento del desempeño en el conjunto de validación. La idea fundamental de este algoritmo es encontrar el punto óptimo donde el modelo generaliza mejor y evitar el sobreajuste. Teniendo en cuenta estas consideraciones, se procede a implementar las diferentes técnicas propuestas en el siguiente análisis.

4.4.1. Modelo LSTM Básico

El primer modelo que se desarrollará se basa en una arquitectura básica de LSTM. Para su implementación, se explorarán diversas configuraciones, considerando diferentes longitudes de ventana temporal, específicamente 30, 60 y 90 días. Además, se evaluarán tres valores distintos para el número de neuronas en la capa LSTM, los cuales han sido utilizados en el área de investigación (16, 32 y 64) (Bodaghi, Owhadi, Nasr, y Sameni, 2023; Gao, Wang, y Zhou, 2021). La variación del número de neuronas en una capa LSTM resulta útil, ya que permite explorar y ajustar la capacidad y complejidad del modelo, adaptándolo a las características específicas de los datos. Esta experimentación permitirá seleccionar la configuración óptima para obtener los mejores resultados de predicción.

Una vez establecida la arquitectura del modelo, se procede a compilarlo utilizando una función de pérdida de error cuadrático medio ('mean_squared_error') y el optimizador Adam. Estas elecciones son fundamentales en la predicción de series de tiempo mediante técnicas de aprendizaje profundo y han sido ampliamente adoptadas por la comunidad científica (Cao, Li, y Li, 2019; Lv, Wang, Gao, y Zhao, 2021). Además de estas configuraciones, se han establecido otros parámetros importantes para el entrenamiento del modelo. El número de épocas de entrenamiento se ha definido en 50, lo que determina la cantidad de veces que el algoritmo recorre el conjunto de datos completo durante el proceso de entrenamiento. Este valor se selecciona cuidadosamente para equilibrar el tiempo de entrenamiento y la capacidad del modelo para aprender patrones y ajustarse a los datos. Por otro lado, se ha establecido un tamaño de lote de 16, lo cual determina la cantidad de muestras de entrenamiento utilizadas en cada paso de actualización de los pesos durante el proceso de entrenamiento. La selección de un tamaño de lote adecuado resulta crucial para garantizar el rendimiento y la estabilidad del entrenamiento del modelo, ya que impacta la velocidad de convergencia y la eficiencia en la utilización de los recursos computacionales disponibles. Para fijar este valor, se ha tomado como referencia la literatura asociada al campo de estudio (Ko y Chang, 2021; Yang, Yan, Huang, y Mei, 2020).

4.4.2. Modelo Stacked LSTM

En este trabajo de fin de grado, se considerará la arquitectura de una LSTM apilada, también conocida como 'Stacked LSTM'. La principal diferencia entre una Stacked LSTM y una LSTM regular se encuentra en su profundidad, como se explicó en el capítulo 3. Una Stacked LSTM consta de múltiples capas de unidades LSTM apiladas una encima de la otra. En este sentido, el proceso de implementación de esta técnica es similar al explicado anteriormente, pero agregando una nueva capa LSTM. Por lo tanto, la configuración de hiperparámetros incluye variaciones en la ventana temporal para la creación de los conjuntos de entrenamiento y prueba (30, 60 y 90 días), así como variaciones en el número de neuronas en las capas implementadas (16, 32 y 64 neuronas).

En este caso, se ha optado por utilizar dos capas en las redes Stacked LSTM. El uso frecuente de dos capas se debe a una combinación de factores relacionados con el rendimiento y la eficiencia del modelo, lo cual ha sido ampliamente adoptado en el campo (Althelaya, El-Alfy, y Mohammed, 2018; Tuan, Nguyen, y Duong, 2022). Una de las razones principales es que al agregar una segunda capa LSTM se logra una representación jerárquica y más compleja de la secuencia de datos, lo que permite al modelo capturar relaciones y dependencias a diferentes niveles de abstracción. Sin embargo, es importante destacar que agregar capas adicionales en las Stacked LSTM no siempre resulta en una mejora significativa en el rendimiento (Tuan et al., 2022). La adición de más capas puede aumentar la complejidad del modelo y, en algunos casos, conducir al sobreajuste, donde el modelo se ajusta demasiado a los datos de entrenamiento y no generaliza bien a nuevos datos. Además, es necesario considerar que entrenar modelos más profundos puede requerir más recursos computacionales y tiempo de entrenamiento, lo que puede limitar su aplicabilidad en entornos con restricciones de recursos, como el nuestro. Bajo estas consideraciones, se tienen en cuenta 18 alternativas, que incluyen las variaciones adicionales del número de neuronas de la nueva capa LSTM implementada en la arquitectura de la red.

4.4.3. Modelo Bi-LSTM

A continuación, se procede al desarrollo de estas alternativas del modelo, mediante la implementación de una arquitectura BiLSTM. Como se discutió en el capítulo 3, este enfoque utiliza dos capas de LSTM: una que procesa la secuencia de datos de izquierda a derecha y otra que lo hace de derecha a izquierda. La metodología de implementación para esta arquitectura es similar a la utilizada para la LSTM básica, con la diferencia de que el modelo incluye una única capa LSTM bidireccional.

En términos generales, el proceso de ajuste de hiperparámetros sigue siendo el mismo, considerando un total de 9 combinaciones de alternativas. Estas combinaciones se forman mediante la variación de las tres opciones de ventana temporal (30, 60 y 90 días) para la

creación de los conjuntos de entrenamiento y prueba, así como la variación de 3 valores para el número de neuronas (16, 32 y 64). Esto nos proporciona un conjunto de opciones para evaluar y determinar la configuración óptima del modelo BiLSTM en función de su rendimiento en la predicción de series temporales. Los valores de épocas y tamaño del batch, se mantienen fijos en los valores definidos para el modelo LSTM básico.

4.4.4. Modelo CNN-LSTM

El último modelo desarrollado se basa en la técnica CNN-LSTM. Como se ha discutido en el capítulo 3, al agregar una capa convolucional antes de la capa LSTM en un modelo CNN-LSTM, se pueden extraer características significativas de la secuencia de entrada. Estas características se utilizan como información adicional para mejorar el rendimiento de la LSTM en la tarea específica de predicción. Además, al introducir la capa convolucional, se puede reducir la cantidad de datos secuenciales que la LSTM debe procesar, lo que resulta en un modelo más eficiente y menos propenso a sobreajustes.

En este modelo particular, que se enfoca en un caso univariante de predicción, se utiliza una capa de convolución unidimensional llamada 'Conv1D'. Basándonos en la revisión de la literatura, se ha fijado el número de filtros ('convolution kernels') en 64 (Mehtab y Sen, 2020; Xu, 2022). En el contexto de una CNN, los filtros se refieren a los pesos aprendidos que se aplican a los datos de entrada durante la operación de convolución. Cuantos más filtros haya, mayor será la capacidad de la red para aprender diferentes características. En la implementación realizada en este trabajo de grado, se emplea una capa 'Conv1D' con 64 filtros, lo que significa que la red aprenderá 64 características o patrones distintos a partir de los datos de entrada. Es importante destacar que se ha incorporado una operación de pooling en la implementación. Se ha añadido una capa 'MaxPooling1D' con un tamaño de pool de 2, lo que reduce la dimensionalidad de la salida de la capa anterior (Mehtab y Sen, 2020). Esta operación se realiza con el fin de disminuir la cantidad de parámetros y mejorar la eficiencia computacional del modelo. Al reducir la dimensionalidad, se logra una representación más compacta de las características extraídas, lo que facilita el procesamiento subsiguiente y reduce la posibilidad de sobreajuste.

4.5. Evaluación de los resultados

En la sección anterior, se ha presentado una descripción detallada de los pasos llevados a cabo en la implementación de las diferentes técnicas. Teniendo en cuenta estas consideraciones, en esta sección, se analizarán los resultados obtenidos, tanto a nivel de cada arquitectura individual como en general, con el objetivo de seleccionar el modelo más preciso.

Como se mencionó en la sección 3.7, se utilizaron varias métricas para evaluar los resultados de todos los modelos, entre las cuales se encuentran: RMSE, MAE y R2. Cada una

de estas métricas proporciona una perspectiva diferente sobre la calidad de las predicciones realizadas por el modelo. El RMSE es especialmente útil cuando se desea penalizar más los errores grandes, ya que los errores al cuadrado amplifican las diferencias. Por otro lado, el MAE es útil para evaluar la precisión promedio de las predicciones sin tener en cuenta la magnitud de los errores. Finalmente, el R2 permite medir la proporción de la varianza en los valores de respuesta que puede ser explicada por el modelo.

Para determinar el modelo con mayor precisión, se han analizado conjuntamente estas métricas. Idealmente, se busca un modelo con valores bajos de RMSE y MAE, lo cual indica errores pequeños y alta precisión en las predicciones realizadas. Además, se busca un valor alto de R2, lo que implica una mayor capacidad del modelo para explicar la variabilidad presente en los datos. En base a estos análisis, se tomará una decisión final respecto al modelo que presente el mejor desempeño y mayor precisión en la predicción de la serie temporal estudiada. Es importante recordar que los resultados presentados, corresponden a una partición 60/40 para los datos de entrenamiento/prueba y a la implementación de una estrategia de detención temprana (*early stopping*) para abordar el problema de sobreajuste que se observó durante la implementación.

4.5.1. Resultados Modelo LSTM

La Tabla 4.3 muestra los resultados obtenidos para la predicción del precio de cierre con un paso adelante (un día adelante) en las diferentes variantes de la arquitectura LSTM del modelo. Se muestran las métricas correspondientes al entrenamiento y la prueba con el propósito de detectar posibles problemas de sobreajuste. El análisis de estas métricas nos permite evaluar el desempeño del modelo en diferentes fases y determinar si existe un desequilibrio entre el ajuste a los datos de entrenamiento y la generalización a nuevos datos de prueba.

Al analizar los resultados presentados en la tabla 4.3, se puede observar que la mayoría de las combinaciones logran un ajuste al modelo superior al 90 % en la predicción del precio de cierre. Sin embargo, es importante señalar que se observa un valor mínimo de R2 en el conjunto de prueba, llegando a ser de 89.20 %. Este resultado se obtuvo específicamente en la combinación que utiliza una ventana temporal de 90 días y 64 neuronas en la capa LSTM. Es relevante mencionar que esta combinación exhibe un comportamiento de sobreajuste, lo que indica que el modelo se ajusta demasiado a los datos de entrenamiento y tiene dificultades para generalizar en nuevos datos del conjunto de prueba. Por otro lado, se destacan los mejores resultados obtenidos en las combinaciones que utilizan una ventana temporal de 30 días en los datos históricos y 16 neuronas en la capa LSTM. Estas combinaciones presentan valores de RMSE, MAE y R2 de 5,7177, 4,49759 y 0,9725. En la Figura 4.7, puede verse el resultado de la predicción para el conjunto de entrenamiento y prueba.

Ventana temporal	Neuronas	Set	RMSE	MAE	R2
30	16	Train	2,7020	2,09132	0,9504
30	16	Test	5,7177	4,49759	0,9725
30	32	Train	2,9014	2,2197	0,9429
30	32	Test	7,9100	6,0370	0,9474
30	64	Train	2,9513	2,2806	0,9419
30	64	Test	7,4989	5,9763	0,9528
60	16	Train	3,6299	2,7805	0,9122
60	16	Test	9,6976	7,4825	0,9136
60	32	Train	2,9513	2,2806	0,9419
60	32	Test	7,3824	5,5323	0,9499
60	64	Train	2,5168	1,9207	0,9578
60	64	Test	5,7884	4,6167	0,9692
90	16	Train	3,5727	2,8000	0,9154
90	16	Test	9,3675	7,1224	0,9112
90	32	Train	3,1775	2,4465	0,9331
90	32	Test	7,3799	5,6226	0,9449
90	64	Train	3,2491	2,4771	0,9300
90	64	Test	10,3328	7,8947	0,8920

Tabla 4.3: Resultado de las variaciones en los parámetros para el modelo LSTM- Un paso adelante

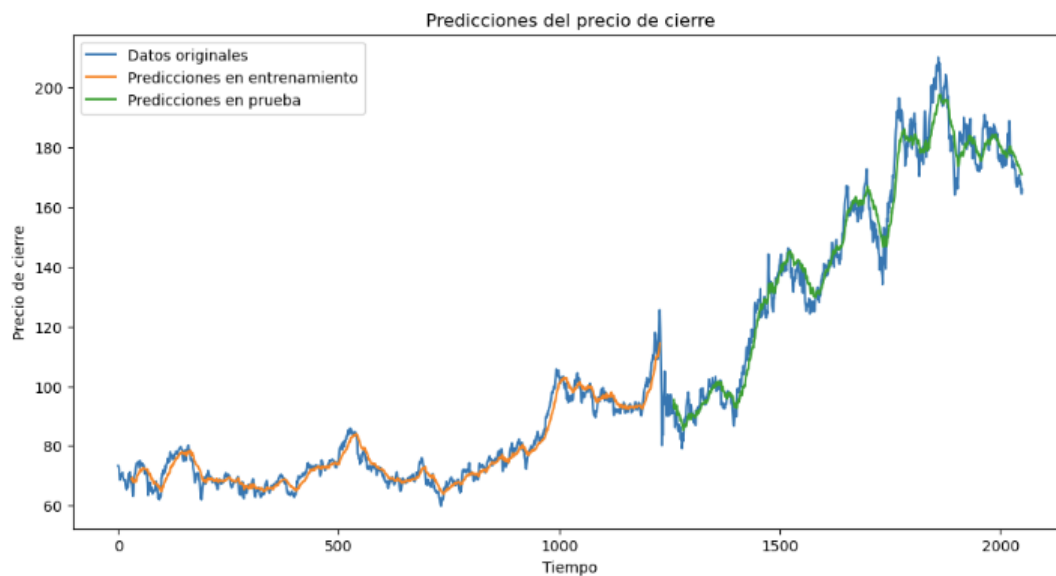


Figura 4.7: Representación del resultado del modelo LSTM 30-16
Elaboración propia

4.5.2. Resultados Modelo Stacked LSTM

Por otra parte, la tabla 4.4 muestra los resultados obtenidos al explorar diversas configuraciones de la ventana temporal y el número de neuronas en las dos capas LSTM que conforman la arquitectura del modelo. En la tabla 4.4, se presentan los resultados obtenidos para diferentes configuraciones de la arquitectura Stacked LSTM. Es notable que la gran mayoría de estas configuraciones superan un valor de R2 del 90 %, lo cual indica un ajuste

satisfactorio del modelo a los datos.

Entre estos resultados, destaca una configuración en particular que muestra el mejor rendimiento en la predicción a corto plazo del precio de cierre de las acciones de Acciona. Esta configuración en específico utiliza una ventana temporal de 90 días y una configuración de 32 neuronas en ambas capas LSTM. Los valores de evaluación para esta configuración son notables, con un RMSE de 5,8492, un MAE de 4,6756 y un R2 de 0,9654.

Ventana temporal	Neuronas Capa 1 - Capa 2	Set	RMSE	MAE	R2
30	16-16	Train	3,2712	2,5435	0,9274
30	16-16	Test	20,1920	16,9912	0,6576
30	16-32	Train	2,8543	2,2029	0,9447
30	16-32	Test	12,093	9,3447	0,8772
30	16-64	Train	2,9436	2,2689	0,9412
30	16-64	Test	6,2559	4,9720	0,9671
30	32-32	Train	2,9814	2,3305	0,9397
30	32-32	Test	16,641	13,6353	0,7674
30	32-64	Train	2,3752	1,8105	0,9617
30	32-64	Test	11,524	8,9658	0,8884
30	64-64	Train	2,4853	1,9253	0,9581
30	64-64	Test	11,408	9,0997	0,8907
60	16-16	Train	2,9006	2,2797	0,9439
60	16-16	Test	7,1803	5,4302	0,9526
60	16-32	Train	2,7777	2,1503	0,9486
60	16-32	Test	10,5309	8,1001	0,8981
60	16-64	Train	2,7572	2,1776	0,9493
60	16-64	Test	8,6425	6,7716	0,9314
60	32-32	Train	2,7033	2,0870	0,9513
60	32-32	Test	10,0383	7,9324	0,9074
60	32-64	Train	3,4349	2,6492	0,9214
60	32-64	Test	8,3542	6,8120	0,9359
60	64-64	Train	2,3788	1,8325	0,9323
60	64-64	Test	8,9905	6,9507	0,9257
90	16-16	Train	3,2202	2,5544	0,9313
90	16-16	Test	6,9653	5,6107	0,9509
90	16-32	Train	2,9209	2,2921	0,9435
90	16-32	Test	12,9307	10,4252	0,8309
90	16-64	Train	2,9692	2,2480	0,9416
90	16-64	Test	6,6638	5,3917	0,9551
90	32-32	Train	2,5086	1,8951	0,9583
90	32-32	Test	5,8492	4,6756	0,9654
90	32-32	Train	2,5097	1,9234	0,9583
90	32-32	Test	7,9088	6,2307	0,9367
90	32-64	Train	2,6538	2,0829	0,9533
90	32-64	Test	6,9582	5,5383	0,9510
90	64-64	Train	2,7589	2,0721	0,9496
90	64-64	Test	18,7329	15,5680	0,6451

Tabla 4.4: Resultado de las alteraciones al modelo Stacked LSTM

La figura 4.8 muestra los resultados de la predicción del modelo obtenido por la arquitectura de 90 días de ventana temporal, y un tamaño de 32 neuronas, en ambas capas.

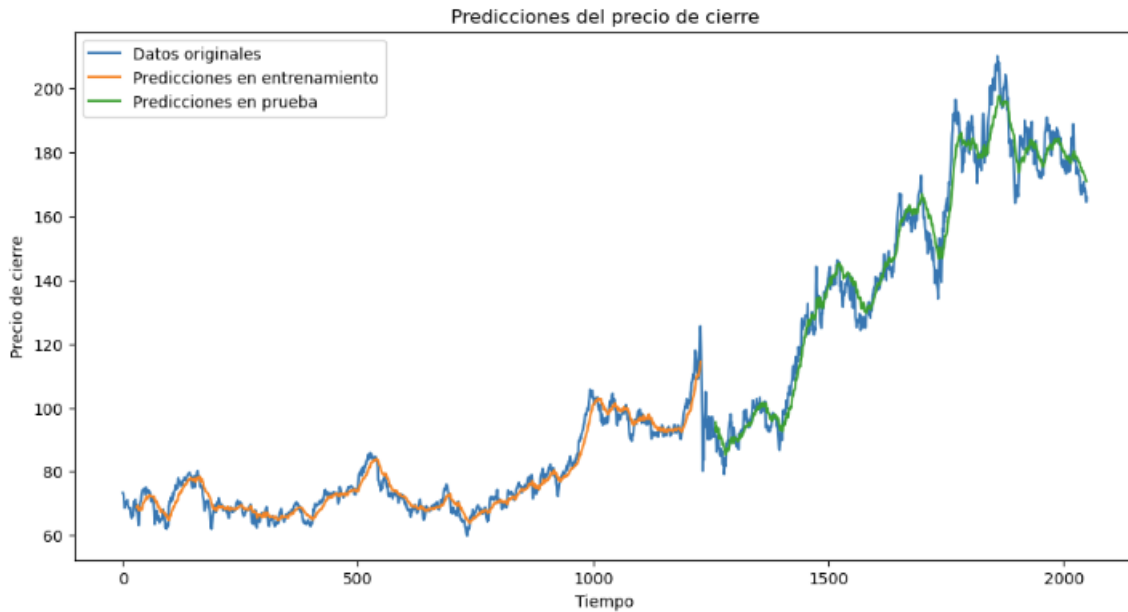


Figura 4.8: Representación del resultado del modelo Stacked LSTM 90-32-32
Elaboración propia

4.5.3. Resultados Modelo BiLSTM

En el caso del modelo BiLSTM, se pueden consultar los resultados obtenidos en la tabla 4.5, donde se muestran los distintos resultados logrados al probar diversas combinaciones de la longitud de la ventana temporal y el número de neuronas en la capa BiLSTM, la cual desempeña un papel clave en la arquitectura de este modelo. Esta tabla refleja los resultados obtenidos en el análisis de estas configuraciones y destaca las métricas asociadas al rendimiento de cada una de ellas.

Al analizar los resultados presentados en la tabla 4.5, se puede observar que las configuraciones asociadas al modelo BiLSTM muestran una precisión ligeramente superior en comparación con las técnicas anteriores. En el conjunto de prueba, la combinación que alcanza el menor valor de R2 es del 81.81 %, la cual se obtiene utilizando una ventana temporal de 60 días para la visualización de los datos históricos y 64 neuronas en la capa BiLSTM.

Por otro lado, la configuración que presenta los mejores resultados, mostrados en la gráfica 4.9, corresponden a una ventana temporal de 60 días y 32 neuronas, con valores de RMSE, MAE y R2 de 5.4678, 4.3296 y 0.9725, respectivamente. Estos resultados demuestran el buen desempeño del modelo BiLSTM en la predicción del valor futuro de los datos analizados en el presente trabajo de fin de grado.

4.5.4. Resultados Modelo CNN-LSTM

En la tabla 4.6 se presentan los resultados obtenidos al evaluar diferentes combinaciones de la longitud de la ventana temporal y el número de neuronas en la capa LSTM del modelo

Ventana temporal	Neuronas	Set	RMSE	MAE	R2
30	16	Train	3,4849	2,6932	0,9176
30	16	Test	7,8385	5,9716	0,9484
30	32	Train	2,9815	2,2651	0,9397
30	32	Test	10,2748	8,1222	0,9113
30	64	Train	2,7219	2,0862	0,9497
30	64	Test	12,0308	9,2446	0,8784
60	16	Train	3,3699	2,6676	0,9243
60	16	Test	7,7541	6,1539	0,9447
60	32	Train	2,3911	1,8496	0,9619
60	32	Test	5,4678	4,3296	0,9725
60	64	Train	2,5132	1,9620	0,9579
60	64	Test	14,0686	10,8542	0,8181
90	16	Train	3,2264	2,5485	0,9310
90	16	Test	6,9513	5,3850	0,9511
90	32	Train	2,5515	1,9886	0,9569
90	32	Test	10,8859	8,6913	0,8801
90	64	Train	2,2647	1,7352	0,9660
90	64	Test	8,7318	6,8075	0,9229

Tabla 4.5: Resultado de las diversas configuraciones del modelo BiLSTM



Figura 4.9: Representación del resultado del modelo BiLSTM 60-32
Elaboración propia

CNN-LSTM. Es importante recordar que esta arquitectura combina una capa convolucional con una capa LSTM para extraer características significativas de la secuencia de entrada y mejorar la capacidad de predicción.

Entre las combinaciones evaluadas observadas en la tabla (4.6), se identifica una configuración que presenta el peor resultado en cuanto al ajuste del modelo, obteniendo un porcentaje de ajuste de tan solo 69.28% al utilizar una ventana temporal de 90 días y 32 neuronas. Es

interesante notar que, en contraste, la predicción para el conjunto de entrenamiento muestra un desempeño considerablemente mejor, con un valor de RMSE de 2.02 y un R2 del 96 %. Esta discrepancia puede ser explicada por la presencia de sobreajuste, que se produce cuando el modelo se ajusta demasiado a los datos de entrenamiento pero no logra generalizar de manera efectiva para realizar predicciones precisas en nuevos datos.

Ventana temporal	Neuronas	Set	RMSE	MAE	R2
30	16	Train	2,5424	1,9499	0,9561
30	16	Test	8,9919	6,9465	0,9321
30	32	Train	2,3590	1,8012	0,9622
30	32	Test	6,9473	5,4144	0,9595
30	64	Train	2,0293	1,5152	0,9721
30	64	Test	4,7384	3,7201	0,9811
60	16	Train	2,3700	1,8364	0,9626
60	16	Test	5,6620	4,3426	0,9705
60	32	Train	2,1232	1,6282	0,9700
60	32	Test	9,0114	7,0274	0,9254
60	34	Train	2,0214	1,5127	0,9728
60	34	Test	5,7652	4,4559	0,9695
90	16	Train	2,3267	1,7877	0,9641
90	16	Test	5,4120	4,3118	0,9710
90	32	Train	2,4012	1,8769	0,9618
90	32	Test	17,4266	14,4430	0,6928
90	64	Train	2,1579	1,6617	0,9691
90	64	Test	8,6274	6,5525	0,9247

Tabla 4.6: Resultado de las alteraciones al modelo CNN-LSTM

Por otro lado, se destaca el modelo con mejor rendimiento dentro de este grupo, cuyos resultados pueden verse en la gráfica 4.10, el cual utiliza una ventana temporal de 30 días y 64 neuronas en la capa LSTM. Esta combinación muestra resultados prometedores con valores de RMSE, MAE y R2 de 4.7384, 3.7201 y 0.9811 respectivamente. Estos resultados indican una menor magnitud de errores en las predicciones y una capacidad más alta para explicar la variabilidad de los datos.

4.5.5. Comparativa general de los modelos implementados

Después de analizar por separado los rendimientos de cada una de las configuraciones en los modelos implementados, donde se varió el tamaño de la ventana y el número de neuronas de las redes neuronales, se realizará una evaluación general de los mejores resultados de cada grupo. El objetivo es determinar cuál de todos los modelos predice con mayor precisión el precio de las acciones de Acciona a corto plazo (un paso adelante).

La tabla 4.7 presenta los resultados del mejor desempeño obtenido en cada una de las configuraciones implementadas para los modelos de Deep Learning propuestos. En todos los casos, se ha realizado una partición del 60 % para el conjunto de entrenamiento y el 40 %



Figura 4.10: Representación del resultado del modelo CNN LSTM 30-64
Elaboración propia

Técnica	Ventana temporal	Neuronas	Set	RMSE	MAE	R2
LSTM	30	16	Train	2,7020	2,09132	0,9504
LSTM	30	16	Test	5,7177	4,49759	0,9725
Stacked LSTM	90	32-32	Train	2,5086	1,8951	0,9583
Stacked LSTM	90	32-32	Test	5,8492	4,6756	0,9654
BiLSTM	60	32	Train	2,3911	1,8496	0,9619
BiLSTM	60	32	Test	5,4678	4,3296	0,9725
CNN-LSTM	30	64	Train	2,0293	1,5152	0,9721
CNN-LSTM	30	64	Test	4,7384	3,7201	0,9811

Tabla 4.7: Conclusión de resultados

para el conjunto de prueba, utilizando 50 épocas y aplicando la estrategia de *early stopping*. Además, se ha utilizado un tamaño de batch de 16 para el entrenamiento de los modelos.

Como se puede observar en la tabla 4.7, todas las técnicas muestran un rendimiento significativamente favorable en la predicción del precio de cotización de las acciones a corto plazo. Sin embargo, hay una técnica que se destaca ligeramente en términos de ajuste. Después de implementar, analizar y comparar 90 modelos diferentes, se concluye que la arquitectura más efectiva es el modelo CNN-LSTM con una ventana temporal de 30 días y 64 neuronas en su capa LSTM. Estos resultados demuestran la capacidad de la técnica CNN-LSTM para capturar patrones temporales en los datos y realizar predicciones precisas en el contexto de las acciones de Acciona.

Capítulo 5

Conclusiones y trabajo futuro

La relevancia de este estudio radica en la necesidad de obtener predicciones precisas sobre los precios de las acciones en el mercado bursátil. Un modelo de predicción a corto plazo del precio de las acciones es sumamente valioso para los posibles inversores, ya que les brinda la capacidad de tomar decisiones informadas en materia de inversión, gestionar el riesgo, identificar oportunidades internacionales, diversificar sus carteras y maximizar sus rendimientos. Asimismo, proporciona información crucial para anticipar los movimientos del mercado, tomar decisiones estratégicas y aumentar las posibilidades de éxito en el ámbito de los valores. Aunque en el pasado se consideraba que las fluctuaciones en los precios de las acciones eran impredecibles, investigaciones recientes han demostrado que la identificación de patrones y tendencias en los movimientos de los precios puede permitir una predicción con cierto grado de precisión. Por lo tanto, este trabajo de fin de grado abordó esta problemática con el objetivo de desarrollar una comparativa de modelos de Deep Learning que contribuyera a mejorar la capacidad de predicción del precio de cierre de las acciones de una compañía en particular. Mediante el análisis y la evaluación de diferentes técnicas y arquitecturas de modelos, se busca proporcionar a los inversores herramientas más eficientes y confiables para tomar decisiones informadas en el mercado bursátil. Por lo tanto, este trabajo de fin de grado abordó esta problemática con el objetivo de desarrollar una comparativa de modelos de Deep Learning que contribuyera a mejorar la capacidad de predicción del precio de cierre de las acciones de una compañía en particular.

Con el objetivo de lograrlo, se llevó a cabo un análisis exhaustivo del estado del arte, que implicó un estudio detallado y una comparación de diversos artículos de investigación que han incorporado técnicas basadas en deep learning para predecir los precios de cierre de acciones a corto plazo. Tras revisar la literatura existente, se observó que la mayoría de las soluciones implementadas para la predicción del precio de las acciones a corto plazo se basan en arquitecturas LSTM o combinaciones de LSTM y CNN. Además, los resultados mostraron que las arquitecturas CNN-LSTM logran resultados ligeramente superiores en cuanto al ajuste de los datos de predicción. En estos modelos, las variables de entrada más utilizadas

incluyen el precio de apertura, cierre, mínimo y máximo del día, junto con el volumen de negociación. Es importante destacar que el precio de cierre se considera la variable más relevante, ya que refleja el valor final de una acción al término del día, proporcionando una medida consolidada del comportamiento del mercado y la evaluación de los inversores. Por tanto, se considera un indicador clave para comprender la tendencia y el rendimiento de las acciones. El análisis de estos trabajos previos y la identificación de las variables más significativas fueron claves para el desarrollo y evaluación de los modelos implementados en este trabajo de fin de grado.

Teniendo en cuenta el contexto actual y la diversidad de estudios realizados en este campo, se tomó la decisión de implementar diversas arquitecturas basadas en unidades LSTM con el objetivo de abordar la problemática de la predicción del precio de las acciones en el mercado bursátil. Se consideraron diferentes enfoques, como la LSTM básica, la Stacked LSTM, la BiLSTM y la CNN-LSTM, cada una de las cuales presenta características y capacidades distintas. La elección de estas arquitecturas se fundamenta en su amplia utilización y reconocimiento en la literatura especializada, así como en su potencial para capturar patrones temporales y captar relaciones complejas en los datos financieros. Al explorar estas diferentes arquitecturas, se buscó evaluar su desempeño y determinar cuál de ellas demostraba un mejor desempeño en cuanto a precisión y capacidad predictiva.

Para la implementación de estas herramientas, se ha seleccionado la empresa Acciona, perteneciente al índice bursátil español IBEX 35, como objeto de análisis del modelo de predicción. Los datos históricos de los precios de las acciones de Acciona fueron obtenidos de Yahoo Finance, una reconocida fuente de información financiera. Se recopilaron datos correspondientes a un período de al menos cinco años, con el objetivo de capturar las tendencias y patrones a largo plazo en los precios de las acciones. Específicamente, los datos abarcan desde el 26 de junio de 2015 hasta el 19 de junio de 2023, y se registraron con una frecuencia diaria. La elección de este período temporal proporciona una cobertura amplia de la evolución de los precios de las acciones de Acciona a lo largo del tiempo.

Posteriormente, se ha realizado un pre-procesamiento y un análisis descriptivo de los datos con el fin de garantizar su calidad y obtener información relevante de la serie temporal. El análisis de los datos históricos de las acciones de Acciona revela que todas las variables cuentan con 2050 observaciones y no presentan valores nulos, lo que asegura una consistencia y completitud en los registros. Al examinar los histogramas de las variables, se confirma que los precios de cierre se concentran principalmente en el rango de 60 a 80 euros, presentando una distribución asimétrica hacia la derecha. Este patrón sugiere una predominancia de valores más altos en comparación con los valores más bajos. Además, al analizar el gráfico de tendencia, se observa un incremento constante en el precio de las acciones de Acciona a lo largo del período analizado, con fluctuaciones significativas debido a eventos como la crisis del Covid-19 y la salida a bolsa de Acciona Energía. Estos eventos se reflejan en momentos de mayor volatilidad y cambios bruscos en el precio. Es interesante destacar que el año 2022

se destaca en el análisis de la frecuencia anual, presentando una dispersión de precios más amplia y alcanzando la máxima cotización. Este fenómeno puede atribuirse a la revalorización de las acciones de Acciona después de la salida a bolsa de Acciona Energía, así como a la recuperación económica tras la pandemia.

Tras realizar el análisis descriptivo de los datos, se procedió al desarrollo de los modelos seleccionados para la comparativa. Sin embargo, durante la etapa inicial de implementación, se detectó un problema de sobreajuste al utilizar una partición del 80 % para el conjunto de entrenamiento y el 20 % para el conjunto de prueba. Con el fin de solucionar este problema, se decidió ampliar la muestra asignando el 40 % para el conjunto de prueba y el 60 % para el conjunto de entrenamiento. Además, se aplicó la técnica conocida como “*early stopping*”, ampliamente utilizada en la literatura de modelos de redes neuronales, con el objetivo de mejorar la capacidad de generalización del modelo y obtener resultados más prometedores en términos de precisión en la predicción. Así, se ajustó el número de épocas y el tamaño del batch en función de los resultados previamente documentados en la literatura, utilizando la nueva partición de datos asignada al 60 % para el conjunto de entrenamiento y al 40 % para el conjunto de prueba. Estas configuraciones se aplicaron de manera consistente a todas las variantes de técnicas de modelos consideradas, las cuales involucraban diferentes ventanas temporales (30, 60 y 90 días) y distintos números de neuronas en la capa LSTM (16, 32 y 64). En total, se desarrollaron y evaluaron 90 modelos diferentes con el objetivo de identificar la combinación que ofreciera la mayor precisión en la predicción del precio de las acciones de Acciona a corto plazo.

Después de ejecutar todos los modelos, se ha podido concluir que una arquitectura de red neuronal CNN-LSTM con una ventana temporal de 30 días y 64 neuronas en su capa LSTM, es la que presenta un mejor ajuste y precisión en la predicción. Esta arquitectura combina las ventajas de las CNN y las LSTM. Estos resultados pueden explicarse debido a que las capas convolucionales de la CNN son altamente efectivas para extraer características y patrones de los datos de entrada, lo cual es especialmente útil para datos de series temporales. Esto permite que la red neuronal aprenda automáticamente las características relevantes de los datos, mejorando así su capacidad de generalización. Además, las capas LSTM de la arquitectura CNN-LSTM son capaces de capturar la dependencia temporal de los datos y modelar secuencias largas de manera efectiva. Esta capacidad es especialmente útil en tareas de predicción de series temporales, donde se requiere analizar la evolución a lo largo del tiempo. Como conclusión, la combinación de estas dos arquitecturas permite aprovechar sus beneficios individuales obteniendo una mayor precisión y capacidad de predicción en el contexto del presente trabajo de grado.

En la línea de trabajo abordada en este proyecto de Fin de Grado, existe la posibilidad de continuar con trabajos futuros que se centren en la mejora del modelo de predicción actual. Esta mejora podría involucrar la exploración de técnicas de aprendizaje automático más avanzadas, como algoritmos de aprendizaje reforzado (reinforcement learning). Además, se

podrían incorporar características adicionales o datos externos relevantes, como noticias financieras o datos macroeconómicos, para enriquecer el modelo y mejorar su capacidad de predicción. Asimismo, se podría realizar una exhaustiva optimización de los hiperparámetros y explorar diferentes estrategias de validación cruzada para evaluar la robustez del modelo en diferentes escenarios. También, partiendo de este mismo modelo, se podría implementar una predicción de múltiples pasos adelante, de 3 o 7 días, en lugar de un solo paso adelante.

Referencias

- Abbasimehr, H., Shabani, M., y Yousefi, M. (2020). An optimized model using lstm network for demand forecasting. *Computers & industrial engineering*, 143, 106435.
- Acciona. (2023). *Proyectos de acciona*. Descargado de https://www.acciona.com/es/proyectos/?_adin=02021864894
- Acciona Yahoo. (2023). *Yahoo finance*. Descargado de <https://es.finance.yahoo.com/quote/ANA.MC?p=ANA.MC&.tsrc=fin-srch>
- Althelaya, K. A., El-Alfy, E.-S. M., y Mohammed, S. (2018). Stock market forecast using multivariate analysis with bidirectional and stacked (lstm, gru). En *2018 21st saudi computer society national computer conference (ncc)* (pp. 1–7).
- Bagheri, A., Peyhani, H. M., y Akbari, M. (2014). Financial forecasting using anfis networks with quantum-behaved particle swarm optimization. *Expert Systems with Applications*, 41(14), 6235–6250.
- Ballman, K. (1998). Forecasting: Methods and applications. *The American Mathematical Monthly*, 105(9), 885.
- Bhanja, S., y Das, A. (2019). Deep learning-based integrated stacked model for the stock market prediction. *Int. J. Eng. Adv. Technol*, 9(1), 5167–5174.
- Bodaghi, F., Owhadi, A., Nasr, A. K., y Sameni, M. K. (2023). A novel cnn-lstm model for predicting the railway sector stock price in the tehran stock exchange. En *2023 9th international conference on web research (icwr)* (pp. 108–114).
- Bülow, S. (2017). The effectiveness of fundamental analysis on value stocks—an analysis of Piotroski’s f-score.
- Cao, J., Li, Z., y Li, J. (2019). Financial time series forecasting model based on ceemdan and lstm. *Physica A: Statistical mechanics and its applications*, 519, 127–139.
- Delteil, C. (2023). *Creating stunning neural network visualizations with chatgpt and plotneuralnet*. Descargado de <https://pub.towardsai.net/creating-stunning-neural-network-visualizations-with-chatgpt-and-plotneuralnet-adab37589e5>
- Fariza, I. (2023). *Acciona ganó 441 millones en 2022, un 33% más, impulsada por su negocio energético*. Descargado de <https://elpais.com/economia/2023-02-27/acciona-gano-441-millones-en-2022-un-33-mas-impulsada>

- por-su-negocio-energetico.html
- Gallego, A. (2023). *Codigo*. Descargado de <https://github.com/AinhoaGallego/TGF.git>
- Gao, Y., Wang, R., y Zhou, E. (2021). Stock prediction based on optimized lstm and gru models. *Scientific Programming*, 2021, 1–8.
- Huang, J., Chai, J., y Cho, S. (2020). Deep learning in finance and banking: A literature review and classification. *Frontiers of Business Research in China*, 14(1), 1–24.
- Jansen, S. (2020). *Machine learning for algorithmic trading: Predictive models to extract signals from market and alternative data for systematic trading strategies with python*. Packt Publishing Ltd.
- Ko, C.-R., y Chang, H.-T. (2021). Lstm-based sentiment analysis for stock price forecast. *PeerJ Computer Science*, 7, e408.
- Kofi, N. I., Adekoya, A. F., y Weyori, B. A. (2020). A systematic review of fundamental and technical analysis of stock market predictions. *The Artificial Intelligence Review*, 53(4), 3007–3057.
- Landi, F., Baraldi, L., Cornia, M., y Cucchiara, R. (2021). Working memory connections for lstm. *Neural Networks*, 144, 334–341.
- Liang, X., Ge, Z., Sun, L., He, M., y Chen, H. (2019). Lstm with wavelet transform based data preprocessing for stock price prediction. *Mathematical Problems in Engineering*, 2019.
- Lu, W., Li, J., Li, Y., Sun, A., y Wang, J. (2020). A cnn-lstm-based model to forecast stock prices. *Complexity*, 2020, 1–10.
- Lv, J., Wang, C., Gao, W., y Zhao, Q. (2021). An economic forecasting method based on the lightgbm-optimized lstm and time-series model. *Computational Intelligence and Neuroscience*, 2021, 1–10.
- Masry, M., y cols. (2017). The impact of technical analysis on stock returns in an emerging capital markets (ecm's) country: Theoretical and empirical study. *International Journal of Economics and Finance*, 9(3), 91–107.
- Mehtab, S., y Sen, J. (2020). Stock price prediction using cnn and lstm-based deep learning models. En *2020 international conference on decision aid sciences and application (dasa)* (pp. 447–453).
- Murphy John, J. (1999). Technical analysis of the financial markets. *New York Institute of Finance*.
- Nabipour, M., Nayyeri, P., Jabani, H., Mosavi, A., y Salwana, E. (2020). Deep learning for stock market prediction. *Entropy*, 22(8), 840.
- O.Nawajana, A. (2020). *Bidirectional lstm model showing the input and output layers*. Descargado de https://www.researchgate.net/figure/Bidirectional-LSTM-model-showing-the-input-and-output-layers-The-red-arrows-represent_fig3_344554659

- Parot, A., Michell, K., y Kristjanpoller, W. D. (2019). Using artificial neural networks to forecast exchange rate, including var-vecm residual analysis and prediction linear combination. *Intelligent Systems in Accounting, Finance and Management*, 26(1), 3–15.
- Patel, J., Shah, S., Thakkar, P., y Kotecha, K. (2015). Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. *Expert systems with applications*, 42(1), 259–268.
- Sethia, A., y Raut, P. (2019). Application of lstm, gru and ica for stock price prediction. En *Information and communication technology for intelligent systems: Proceedings of ictis 2018, volume 2* (pp. 479–487).
- Song, H. (2017). *Review of time series analysis and its applications with r examples*, by robert h. shumway & david s. stoffer: New york, ny: Springer, 2011, 596 pp., 99,00(hardcover). Taylor&Francis.
- Swathi, T., Kasiviswanath, N., y Rao, A. A. (2022). An optimal deep learning-based lstm for stock price prediction using twitter sentiment analysis. *Applied Intelligence*, 52(12), 13675–13688.
- Tuan, N. T., Nguyen, T. H., y Duong, T. T. H. (2022). Stock price prediction in vietnam using stacked lstm. En *Intelligence of things: Technologies and applications: The first international conference on intelligence of things (iciot 2022), hanoi, vietnam, august 17–19, 2022, proceedings* (pp. 246–255).
- Wang, H., Wang, J., Cao, L., Li, Y., Sun, Q., y Wang, J. (2021). A stock closing price prediction model based on cnn-bislstm. *Complexity*, 2021, 1–12.
- Wang, J.-J., Wang, J.-Z., Zhang, Z.-G., y Guo, S.-P. (2012). Stock index forecasting based on a hybrid model. *Omega*, 40(6), 758–766.
- Wu, J. M.-T., Li, Z., Herencsar, N., Vo, B., y Lin, J. C.-W. (2021). A graph-based cnn-lstm stock price prediction algorithm with leading indicators. *Multimedia Systems*, 1–20.
- Wu, X.-x., y Liu, J.-g. (2009). A new early stopping algorithm for improving neural network generalization. En *2009 second international conference on intelligent computation technology and automation* (Vol. 1, pp. 15–18).
- Xu, W. (2022). Stock price prediction based on cnn-lstm model in the pytorch environment. En *2022 2nd international conference on economic development and business culture (icedbc 2022)* (pp. 1272–1276).
- Yang, Z., Yan, W., Huang, X., y Mei, L. (2020). Adaptive temporal-frequency network for time-series forecasting. *IEEE Transactions on Knowledge and Data Engineering*, 34(4), 1576–1587.