Facultad de Ciencias Económicas y Empresariales

Influencia del Learning Rate en el Desempeño de Agentes de Deep Reinforcement Learning para Estrategias de Gestión de Carteras

Autor: Román Martín Gallego
Director: Eduardo César Garrido

Clave: 201900804

MADRID | Abril 2024

Resumen

Este Trabajo de Fin de Grado aborda la influencia de la tasa de aprendizaje en la eficacia de los agentes de Deep Reinforcement Learning (DRL) aplicados a estrategias de gestión de carteras financieras. Dada la volatilidad y la complejidad inherente a los mercados financieros, este estudio se enfoca en entender el impacto del learning rate en un modelo de DRL para mejorar las decisiones en la gestión de carteras.

La investigación emplea el algoritmo de Proximal Policy Optimization (PPO), explorando cómo diferentes configuraciones del learning rate afectan la capacidad de los modelos de DRL para adaptarse a las dinámicas del mercado y maximizar la rentabilidad ajustada al riesgo, medida a través del ratio de Sharpe. Se propone una hipótesis principal que postula que un learning rate incrementado facilitará una adaptación más efectiva y rápida a las condiciones cambiantes del mercado, superando en rendimiento a la configuración predeterminada en Stable Baselines 3. El estudio metodológico incluye un diseño experimental donde se compara la performance de múltiples agentes con tasas de aprendizaje variadas, utilizando datos históricos del mercado para simular escenarios de inversión. Los resultados esperan demostrar que ajustes precisos en el learning rate pueden ofrecer ventajas significativas en términos de eficiencia y efectividad de las estrategias de inversión automatizadas.

Este trabajo contribuye al campo de las finanzas cuantitativas al ofrecer una comprensión más profunda de cómo los parámetros de aprendizaje por refuerzo, más concretamente, el learning rate, influyen en el desempeño de los algoritmos de DRL en finanzas.

Abstract

This thesis addresses the influence of the learning rate on the effectiveness of Deep Reinforcement Learning (DRL) agents applied to financial portfolio management strategies. Given the volatility and inherent complexity of financial markets, this study focuses on understanding the impact of the learning rate on a DRL model to improve decisions in portfolio management.

The research utilizes the Proximal Policy Optimization (PPO) algorithm, exploring how different settings of the learning rate affect the DRL models' ability to adapt to market dynamics and maximize risk-adjusted returns, measured through the Sharpe ratio. A main hypothesis proposes that an increased learning rate will facilitate a more effective and rapid adaptation to changing market conditions, outperforming the default setting in Stable Baselines 3. The methodological study includes an experimental design where the performance of multiple agents with varied learning rates is compared, using historical market data to simulate investment scenarios. The results are expected to demonstrate that precise adjustments in the learning rate can offer significant advantages in terms of efficiency and effectiveness of automated investment strategies.

This work contributes to the field of quantitative finance by offering a deeper understanding of how reinforcement learning parameters, specifically the learning rate, influence the performance of DRL algorithms in finance.

Índice general

Índice de ilustraciones

Índice de ecuaciones

Capítulo 1

Introducción

En las últimas décadas, el campo de las finanzas ha experimentado una revolución significativa debido a la intersección con la tecnología, particularmente en lo que respecta al aprendizaje automático (Sánchez, 2022). Este avance tecnológico ha propiciado una evolución notable en las metodologías y enfoques utilizados en el trading financiero, un área que ha sido influenciada notablemente por estos cambios (Dixon, Halperin, & Bilokon, 2020). El presente proyecto de fin de grado se inscribe en esta evolución, centrando su atención en una de las áreas más prometedoras de esta intersección: la aplicación del aprendizaje por refuerzo (AR) (González Oviedo, 2023) en operaciones de mercado, con un enfoque específico en el desarrollo de una red neuronal para la toma de decisiones de trading.

Históricamente, las finanzas cuantitativas han sido dominadas por modelos estadísticos y técnicas de aprendizaje automático, principalmente métodos de aprendizaje supervisado. Estudios pioneros, como los de Liu, Yang, Gao, y Wang (2021) y Joshi (2003), han mostrado el potencial de las redes neuronales y las máquinas de vectores de soporte en la predicción del rendimiento de los activos financieros. Sin embargo, estas metodologías, aunque efectivas, a menudo se encontraban limitadas por su enfoque en dos pasos: primero, la predicción de movimientos de mercado y, segundo, la aplicación de estas predicciones en estrategias de trading.

El AR emerge como un paradigma innovador y disruptivo, prometiendo una integración más holística y dinámica de la predicción y ejecución en el trading financiero. Este enfoque se diferencia de los métodos tradicionales al enfocarse en aprender políticas de decisión óptimas a través de la interacción con el entorno, maximizando así una señal de recompensa numérica. Esta característica del AR le confiere el potencial para superar las limitaciones de los modelos predictivos tradicionales (Dixon, Halperin, & Bilokon, 2020).

El proyecto se basa en el diseño e implementación de una red neuronal avanzada. Inicia con el uso de técnicas de aprendizaje por refuerzo, un enfoque que permite a los modelos aprender a tomar decisiones optimizadas a través de la experiencia. Esta base se extiende luego al empleo de técnicas de Deep Reinforcement Learning (Meyer et al., 2019), profundizando el aprendizaje por refuerzo con redes neuronales profundas, para operar de manera más efectiva en el mercado de valores. El sistema propuesto no solo busca predecir movimientos del mercado, sino también aprender a ejecutar operaciones de manera óptima. Se enfoca en un manejo eficiente de activos en la gestión de carteras, un proceso que va más allá de la simple compra o venta de productos. Este enfoque integral busca gestionar una variedad de activos teniendo en cuenta los riesgos asociados para lograr un balance entre rentabilidad y riesgo. Con este proyecto, se pretende abordar la complejidad del trading financiero empleando algoritmos de Deep Reinforced

Learning que se adaptan y aprenden de los cambios del mercado, maximizando los beneficios mientras gestionan los riesgos de manera eficiente (Ruiz Rueda, 2021).

La motivación para la aplicación del AR en el trading financiero es variada y significativa. Desde el deseo de mejorar el rendimiento del trading hasta la necesidad de comprender mejor las complejidades del mercado financiero, el AR se presenta como una solución prometedora. Los avances tecnológicos y el aumento en la capacidad computacional han abierto nuevas oportunidades para explorar modelos más complejos y sofisticados de AR, que no solo mejoran el rendimiento del trading, sino que también proporcionan una comprensión más profunda de los mercados financieros a través de la identificación de señales sutiles y complejas interacciones (Joshi, 2022; Zhu, 2022; Tsantekidis et al., 2020; Fiorini & Fiorini, 2021; Ahmed, Ghoneim, & Saleh, 2020).

Para abordar estos desafíos y motivaciones, se propone un sistema de trading innovador utilizando DRL, basado en la obra de Eduardo C. Garrido Merchán (2023). Este sistema se centrará en una arquitectura de red neuronal profunda, aplicando técnicas como Q-Learning, Deep Q-Learning Networks (DQN), y algoritmos avanzados como A2C, TRPO y PPO. Su distintivo radica en la capacidad de adaptarse continuamente a las condiciones cambiantes del mercado, optimizando la gestión de carteras y riesgos. Para garantizar su eficacia y viabilidad práctica, el sistema se someterá a pruebas exhaustivas en entornos que simulen las condiciones reales del mercado.

Este Trabajo de Fin de Grado está organizado en cinco capítulos. Tras esta introducción, el segundo capítulo ofrece un resumen de la investigación existente sobre los algoritmos de DRL y sus aplicaciones en el ámbito financiero. El tercer capítulo detalla la definición del proyecto, presentando la hipótesis y los objetivos, así como las asunciones y limitaciones encontradas durante la investigación. El cuarto capítulo se dedica a explorar el marco teórico, abarcando todos los conceptos necesarios para comprender y ejecutar el experimento propuesto. Finalmente, el quinto capítulo conduce la transición del marco teórico a la aplicación práctica a través de experimentos, donde se discuten el análisis de datos, la metodología aplicada y la evaluación de los resultados obtenidos.

Capítulo 2

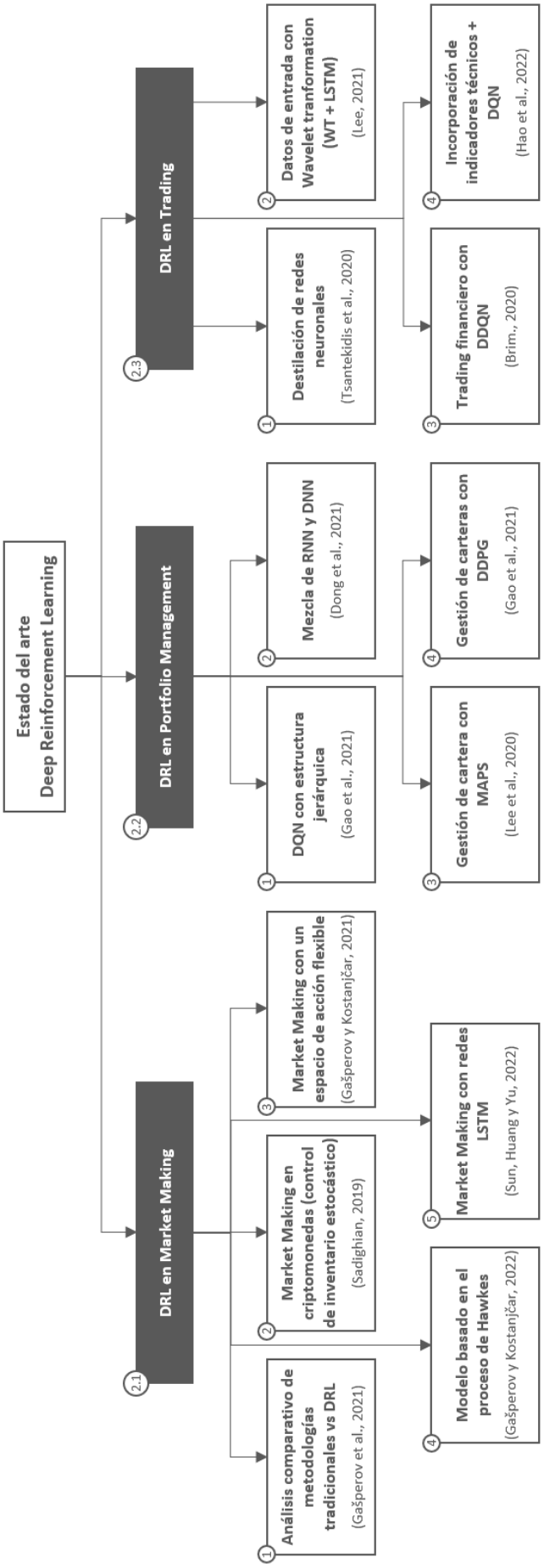
Estado del arte

El Deep Reinforcement Learning (DRL) representa una frontera fascinante en el ámbito de las finanzas, marcando un cambio significativo desde los métodos tradicionales basados en análisis estadísticos y teorías económicas clásicas. Su importancia radica en su capacidad para aprender y adaptarse a través de la interacción con el entorno, lo que lo hace especialmente adecuado para los mercados financieros, caracterizados por su constante cambio y complejidad (Garrido, 2024).

A diferencia de las técnicas analíticas que a menudo están diseñadas para escenarios específicos, el DRL permite un enfoque más dinámico y flexible. A través de algoritmos avanzados, estos sistemas pueden capturar y responder a las dinámicas del mercado en tiempo real, ofreciendo una herramienta poderosa para la toma de decisiones y la gestión de activos. Esta evolución hacia el aprendizaje automático en las finanzas está alcanzando a los modelos clásicos (Merchán, 2023).

Con este fondo, exploraremos en detalle los avances y aplicaciones del DRL en Finanzas. Examinaremos estudios clave que han marcado hitos en este campo, discutiendo sus metodologías, innovaciones y las implicaciones de sus hallazgos para el futuro de las finanzas. La revisión exhaustiva de la literatura previa es fundamental para comprender el estado actual de la investigación en DRL aplicado a finanzas, además de para identificar brechas de conocimiento y oportunidades para tener en cuenta en este proyecto. Esta comprensión detallada nos permitirá establecer un marco sólido para nuestro análisis y contribuir significativamente al cuerpo de conocimiento existente, asegurando que nuestro trabajo esté firmemente arraigado en la vanguardia de la investigación financiera contemporánea.

En vista de los objetivos y restricciones específicos de este Trabajo de Fin de Grado, es esencial delimitar el alcance de nuestra revisión literaria. Por ello, nos centraremos primordialmente en el estudio de Deep Reinforcement Learning aplicado a tres áreas críticas: Market Making, Portfolio Management y Trading. Esta concentración temática no es arbitraria; refleja una estrategia deliberada para abordar los aspectos más relevantes y actuales del DRL en el contexto financiero. Al profundizar en estas áreas, no solo ganaremos una comprensión integral de las técnicas más avanzadas y sus aplicaciones prácticas, sino que también adquiriremos los conocimientos esenciales para el diseño y desarrollo de nuestro agente automatizado. Este agente, concebido para operar en los mercados financieros, será el pilar de nuestra investigación práctica y la culminación de nuestro análisis teórico. En el siguiente esquema se representa la estructura de esta sección para facilitar el seguimiento del estado del arte enfocado al DRL.


Estado del arte
Deep Reinforcement Learning
2.1
DRL en Market Making
1
Análisis comparativo de metodologías tradicionales vs DRL
(Gašperov et al., 2021)
2
Market Making en criptomonedas (control de inventario estocástico)
(Sadighian, 2019)
3
Market Making con un espacio de acción flexible
(Gašperov y Kostanjčar, 2021)
4
Modelo basado en el proceso de Hawkes
(Gašperov y Kostanjčar, 2022)
5
Market Making con redes LSTM
(Sun, Huang y Yu, 2022)
2.2
DRL en Portfolio Management
1
DQN con estructura jerárquica
(Gao et al., 2021)
2
Mezcla de RNN y DNN
(Dong et al., 2021)
3
Gestión de cartera con MAPS
(Lee et al., 2020)
4
Gestión de carteras con DDPG
(Gao et al., 2021)
2.3
DRL en Trading
1
Destilación de redes neuronales
(Tsantekidis et al., 2020)
2
Datos de entrada con Wavelet tranformation (WT + LSTM)
(Lee, 2021)
3
Trading financiero con DDQN
(Brim., 2020)
4
Incorporación de indicadores técnicos + DQN
(Hao et al., 2022)

2.1. Deep Reinforcement Learning en Market Making

La práctica del "Market Making" es un componente esencial en los mercados financieros, desempeñando un papel vital en la provisión de liquidez y en la facilitación de la compra y venta de activos financieros (Debie et al., 2023). Según Qi y Ventre (2022), los market makers son fundamentales para la salud de los mercados financieros, asegurando la liquidez del mercado y permitiendo su funcionamiento eficiente. Erokhin y Lukashenko (2022) destacan la función de los market makers en comprar y vender activos financieros a precios cotizados públicamente, obteniendo beneficios a través del "spread". Esta actividad continua contribuye significativamente a la eficiencia y fluidez del mercado. La incorporación del Aprendizaje Profundo por Refuerzo en el market making, como señala Debie et al. (2023), representa una evolución significativa en este campo, ofreciendo una mayor adaptación y eficiencia en la formación de mercados.

El estudio de la literatura previa sobre la aplicación de Deep Reinforcement Learning en market making es fundamental para nuestro proyecto de desarrollar un agente automatizado para operar en los mercados financieros. Esta revisión es crucial porque en el market making, el DRL se ha enfocado en aspectos esenciales como la estructura de precios, los spreads, y la adaptabilidad a condiciones inciertas del mercado. Estos temas son directamente relevantes para nuestro agente de trading, ya que proporcionan insights valiosos sobre cómo manejar la volatilidad del mercado, optimizar decisiones de compraventa y adaptarse a las dinámicas cambiantes del mercado (Sun et al., 2022).

A continuación, presentaremos una selección de estudios relevantes en el campo del market making, enfocándonos en sus hallazgos clave y las técnicas de Deep Reinforcement Learning (DRL) que emplearon. Al examinar estos estudios, destacaremos tanto las soluciones propuestas como las metodologías utilizadas, lo cual es esencial para informar y guiar el desarrollo de nuestro propio agente de trading automatizado.

El estudio integral de **Gašperov, Begušić, Šimović y Kostanjčar (2021)** proporciona un análisis comparativo con metodologías tradicionales y destacando la superioridad y versatilidad del DRL en diversos escenarios del mercado financiero. El estudio se enfocó en el uso del Deep Reinforcement Learning en la formación de mercados financieros, proveyendo un análisis comparativo con las metodologías tradicionales. Esta investigación se distingue por su enfoque exhaustivo y comparativo, evaluando las técnicas de DRL frente a los enfoques analíticos convencionales que han sido la norma en el sector financiero.

Inicialmente, el equipo revisó detalladamente los modelos tradicionales de formación de mercados, basados en análisis estadísticos y teorías económicas clásicas. Aunque estos modelos han sido históricamente útiles, presentan limitaciones en su capacidad para adaptarse a las dinámicas cambiantes del mercado y capturar complejidades no lineales, una brecha cada vez más evidente en los entornos financieros modernos.

El estudio entonces se centró en cómo el DRL, como una forma avanzada de aprendizaje automático, ofrece una alternativa más flexible y dinámica a estos modelos tradicionales. Los algoritmos de DRL se caracterizan por aprender y adaptarse continuamente a través de la interacción con el entorno, lo que los hace particularmente adecuados para los mercados financieros, donde las condiciones están en constante evolución y son impredecibles.

Una de las contribuciones más significativas de este estudio fue demostrar la superioridad del DRL sobre los métodos tradicionales. Los autores encontraron que los modelos basados en DRL superaban en rendimiento a los enfoques convencionales, especialmente en términos de rentabilidad y capacidad para manejar situaciones de mercado imprevistas y volátiles. La eficacia

del DRL se atribuye a su habilidad inherente para aprender de la experiencia y ajustar estrategias en tiempo real, una ventaja clave en la toma de decisiones financieras.

En **2019, Sadighian** llevó a cabo un estudio destacado en el ámbito del Deep Reinforcement Learning (DRL) aplicado a los mercados de criptomonedas, un sector conocido por su alta volatilidad y complejidad. La investigación de Sadighian es de gran relevancia, principalmente por su enfoque en los desafíos asociados con el control de inventario estocástico en los libros de órdenes limitadas, un elemento crítico en la operación de los mercados financieros.

Para abordar estos desafíos, Sadighian implementó dos algoritmos avanzados de gradiente de políticas, una técnica de aprendizaje por refuerzo que optimiza los parámetros de un algoritmo con el fin de maximizar la utilidad en las decisiones de trading. Estos algoritmos se aplicaron dentro de un entorno simulado, diseñado para replicar el espacio de observación utilizando datos reales de libros de órdenes limitadas y estadísticas de llegada de órdenes. Este enfoque innovador permitió que el sistema de DRL se adaptara y respondiera de manera efectiva a las condiciones cambiantes del mercado en tiempo real.

La evaluación del desempeño de este sistema se realizó analizando los retornos comerciales diarios y el promedio de estos, para cada combinación de agente y función de recompensa. Los resultados obtenidos en el estudio de Sadighian demostraron la capacidad del DRL para manejar eficientemente los desafíos del control de inventario en mercados tan impredecibles y volátiles como el de las criptomonedas.

Continuando con la evolución del DRL en los mercados financieros, **el estudio de Gašperov y Kostanjčar en 2021** lleva la aplicación del DRL un paso adelante, centrando su atención en la integración de señales predictivas y el desarrollo de un espacio de acción más flexible y una función de recompensa innovadora.

Esta investigación pionera, marcó un antes y un después en la aplicación del DRL en el mercado financiero, introduciendo un enfoque novedoso que se aparta de las técnicas analíticas y de aprendizaje automático tradicionales. Su principal innovación radica en la integración única de señales predictivas dentro del marco del DRL, permitiendo que el modelo capturara y utilizara información de mercado de manera más efectiva y dinámica, y se adaptara eficientemente a las condiciones cambiantes del mercado.

Un aspecto distintivo del trabajo de Gašperov y Kostanjčar es el desarrollo de un espacio de acción y una función de recompensa novedosos. En el contexto del DRL, el espacio de acción es crucial ya que define el conjunto de posibles acciones que el modelo puede elegir. Por su parte, la función de recompensa evalúa la efectividad de estas acciones. El diseño de estos componentes es fundamental para reflejar con precisión las complejidades y dinámicas del mercado financiero real. El espacio de acción propuesto por Gašperov y Kostanjčar ofrecía una gama más amplia y flexible de decisiones de inversión, esencial para la operativa en mercados financieros caracterizados por su constante cambio. Esta flexibilidad permitió al modelo de DRL adaptarse rápidamente a nuevas condiciones de mercado, optimizando decisiones en tiempo real.

Por otro lado, la función de recompensa que idearon estaba diseñada para alinear el aprendizaje del modelo con los objetivos a largo plazo del mercado. Esta función no solo incentivaba el rendimiento a corto plazo, sino que también consideraba la estabilidad y sostenibilidad a largo plazo, reflejando una comprensión más profunda y estratégica de los mercados financieros.

Avanzando hacia aplicaciones más complejas y realistas del DRL en el mercado financiero, el estudio posterior de **Gašperov y Kostanjčar en 2022** explora el uso de un modelo basado en el proceso de Hawkes, es un tipo de modelo matemático utilizado para predecir y

analizar eventos que se auto catalizan en el tiempo, es decir, donde la ocurrencia de un evento aumenta la probabilidad de que ocurran eventos futuros. (Han, Ma, Wang, Günnemann, & Tresp, 2020).

Este estudio avanzado se enfocó en el desarrollo y entrenamiento de un controlador basado en DRL diseñado para operar en un entorno de simulación de un libro de órdenes limitadas, un componente crítico en los mercados financieros donde las órdenes de compra y venta se organizan a diferentes niveles de precios. Esta investigación se sumerge en la naturaleza compleja y dinámica de los mercados financieros, ofreciendo una oportunidad única para probar la eficacia de los algoritmos de DRL bajo condiciones realistas del mercado.

La innovación clave de este estudio radica en cómo el controlador de DRL fue entrenado para aprender y adaptarse a las dinámicas cambiantes del mercado. Utilizando datos del libro de órdenes, el controlador tomaba decisiones estratégicas de compra y venta, enfocando su entrenamiento en maximizar la rentabilidad y minimizar los costos, incluso en escenarios de alta volatilidad y elevados costos de transacción.

Los resultados de la investigación de Gašperov y Kostanjčar demostraron que el controlador de DRL no solo era capaz de operar eficientemente en un entorno de mercado complejo, sino que también superaba varios benchmarks establecidos en el ámbito de la formación de mercados. Esta superioridad se mantuvo incluso en contextos con altos costos de transacción, destacando la eficacia del modelo en situaciones desafiantes comunes en las estrategias de trading. Lo más destacable del estudio fue la robustez y adaptabilidad del modelo. El controlador mostró una habilidad notable para manejar tano las condiciones normales del mercado, como también las situaciones inusuales y volátiles.

Siguiendo la trayectoria de innovación en el DRL, el estudio de **Sun, Huang y Yu en 2022** introduce una nueva dimensión en la aplicación del Deep Reinforcement Learning en la formación de mercados financieros, especialmente por introducir el uso de redes LSTM (Long Short-Term Memory) en un modelo de DRL para analizar los libros de órdenes limitadas. Este enfoque innovador marcó un hito en la capacidad del DRL para procesar y comprender las dinámicas de los mercados financieros.

Las redes LSTM, que son una forma avanzada de redes neuronales recurrentes, se destacan por su habilidad en el procesamiento de secuencias temporales y la captura de dependencias a largo plazo. Integrar estas redes en un modelo de DRL permitió al sistema analizar y predecir patrones de compra y venta a lo largo del tiempo, una capacidad crucial para comprender y reaccionar al mercado financiero.

El modelo propuesto por Sun, Huang y Yu es notable por su habilidad para procesar y aprender tanto de datos históricos como de datos en tiempo real del libro de órdenes. Esta capacidad de análisis profundo y la habilidad para predecir cambios en el mercado permitieron al modelo tomar decisiones informadas sobre la compra y venta de activos financieros. La implementación de las redes LSTM permitió al modelo superar las limitaciones de los modelos tradicionales, identificando patrones complejos y tendencias ocultas en los datos del mercado.

Los resultados del estudio de Sun, Huang y Yu demostraron que este enfoque mejoraba significativamente el rendimiento en comparación con los modelos de formación de mercados existentes. Generaba predicciones más precisas, y también mostraba una mayor adaptabilidad a las condiciones cambiantes del mercado. Esta adaptabilidad es esencial para mantener la rentabilidad y minimizar los riesgos en un entorno de mercado altamente volátil.

En el **análisis comparativo** de los anteriores estudios aplicados a la formación de mercados financieros, es esencial considerar tres aspectos clave: eficacia en diversos entornos de

mercado, adaptabilidad a cambios y volatilidad del mercado, y la aplicabilidad y versatilidad de los modelos.

Por un lado, las investigaciones realizadas por Gašperov y Kostanjčar en 2021 y 2022 han demostrado la eficacia del DRL en capturar las dinámicas complejas del mercado, especialmente a través de su enfoque innovador en la creación del espacio de acción y la función de recompensa. Estos avances han ampliado significativamente la comprensión de las operaciones de mercado, marcando un cambio paradigmático en la formación de mercados financieros. Por otro lado, Sadighian (2019) enfocó su estudio en los mercados de criptomonedas, caracterizados por su alta volatilidad, donde demostró la efectividad del DRL utilizando algoritmos de gradiente de políticas. Esta aplicación específica resalta la versatilidad del DRL en entornos de mercado distintos y complejos. Además, Sun, Huang y Yu (2022) incorporaron redes LSTM en modelos de DRL, potenciando la capacidad de estos sistemas para procesar y aprender de datos históricos y en tiempo real, lo que mejora notablemente la precisión y adaptabilidad en las decisiones de mercado.

Por otro lado, estos estudios también explican como la capacidad de adaptarse a condiciones cambiantes y volátiles es crucial en los mercados financieros. La investigación avanzada de Gašperov y Kostanjčar (2022) subraya la capacidad del DRL para adaptarse a estos cambios, incluso en contextos de alta volatilidad y costos de transacción elevados. De manera similar, tanto Sadighian (2019) como Sun, Huang y Yu (2022) demostraron que sus modelos de DRL podían adaptarse eficazmente a las dinámicas cambiantes del mercado, respondiendo adecuadamente a situaciones imprevistas. Estas características destacan la relevancia del DRL en entornos de mercado en constante evolución.

Por último, cabe destacar el estudio integral de Gašperov, Begušić, Šimović y Kostanjčar (2021) donde resalta la versatilidad del DRL en comparación con los métodos analíticos tradicionales. Este estudio demuestra que el DRL es aplicable a una amplia gama de situaciones de mercado, ofreciendo una herramienta flexible y dinámica que supera las limitaciones de los enfoques convencionales.

2.2. Deep Reinforcement Learning en Portfolio Management

La gestión de carteras de inversión, o Portfolio Management, constituye un proceso financiero esencial donde la asignación de activos en una cartera de inversiones busca maximizar los retornos y minimizar los riesgos (Garrido, 2023). Este proceso requiere una selección y gestión cuidadosa de diversos activos, tales como acciones y bonos entre otros, basándose en los objetivos de inversión, el horizonte temporal y la tolerancia al riesgo del inversor.

El desarrollo del Aprendizaje Profundo por Refuerzo (Deep Reinforcement Learning) ha marcado un hito en este ámbito, introduciendo enfoques innovadores para optimizar las estrategias de inversión (Giménez, 2012). Investigaciones como las de Lucarelli y Borrotti (2020) y Yuan Gao et al. (2021) han evidenciado la capacidad del DRL para adaptarse efectivamente a mercados dinámicos y tomar decisiones informadas en escenarios inciertos. Gracias a su naturaleza como subcampo del aprendizaje automático, el DRL permite a los sistemas aprender y mejorar su desempeño interactuando con el entorno, una habilidad particularmente valiosa en los mercados financieros, caracterizados por su rápida evolución.

Investigaciones adicionales, como las realizadas por Dong et al. (2021) y Fazli et al. (2022), resaltan la aplicación del DRL en la optimización de la asignación de activos y la gestión de riesgos, aspectos críticos en un entorno marcado por la alta volatilidad y la complejidad de los

datos. De manera similar, estudios como los de Lee et al. (2020) y Gao et al. (2021) demuestran cómo el DRL es capaz de manejar grandes volúmenes de datos y variables complejas, facilitando la toma de decisiones más informadas y oportunas.

En la siguiente sección de este trabajo, profundizaremos en los estudios más significativos del DRL aplicado a la gestión de carteras. Nuestro enfoque se centrará en los hallazgos clave y en las técnicas específicas de DRL empleadas, lo cual es crucial para orientar el desarrollo de nuestro agente que operará de manera automática en los mercados financieros.

El estudio de **Gao et al. (2021)** incorpora el Deep Q-Network (DQN) en la gestión de carteras de inversión. DQN es una avanzada técnica de aprendizaje por refuerzo que combina el tradicional Q-Learning con redes neuronales profundas. En el Q-Learning, el agente aprende una función de valor (denominada Q) que estima el valor esperado de las recompensas futuras para cada acción en un estado dado. Esta función de valor es esencial para que el agente tome decisiones informadas (Gao et al., 2020).

El uso de redes neuronales en DQN permite al sistema aproximar la función de valor Q, facilitando el manejo de situaciones con un gran número de estados y acciones, como es el caso en la gestión de carteras. En este contexto, el agente (el algoritmo de DQN) aprende a tomar decisiones de inversión óptimas basadas en datos históricos y actuales del mercado, como los precios de las acciones, buscando maximizar la rentabilidad de la cartera y gestionar el riesgo (Gao et al., 2020).

Lo que distingue al modelo de Gao et al. es su enfoque jerárquico. En lugar de utilizar un único DQN para toda la cartera, el modelo divide la cartera en partes más manejables, asignando un DQN a cada segmento. Esta estructura jerárquica permite una mayor flexibilidad y eficiencia, particularmente útil cuando se maneja una amplia gama de activos. Además, esta división puede reducir significativamente los costos de transacción, un factor crítico para la rentabilidad en entornos donde las comisiones pueden impactar de manera considerable.

Los experimentos de Gao et al. utilizando series temporales de acciones en diferentes períodos demuestran que su estrategia basada en DQN jerárquico supera a otras diez estrategias en rentabilidad. Además, al evaluar el riesgo a través del Ratio de Sharpe y el Máximo Drawdown, se reveló que la estrategia asociada con este modelo jerárquico de DQN presentaba el menor riesgo en comparación con las demás.

El estudio innovador de **Dong, Huang, Ma y Qian (2021)** combina técnicas avanzadas de aprendizaje profundo por refuerzo con la gestión de carteras en los mercados de valores. Utilizando una mezcla de Redes Neuronales Recurrentes (RNN) y Redes Neuronales Profundas (DNN), el modelo busca capturar eficientemente la dinámica de los mercados financieros y tomar decisiones estratégicas de inversión.

Las RNN se aplican para analizar series temporales de datos financieros, como los precios de las acciones y los indicadores económicos. Su capacidad para procesar y "recordar" información anterior las hace especialmente adecuadas para predecir tendencias del mercado y precios futuros, considerando la evolución temporal de los datos. Esta característica es crucial para modelar y entender la secuencia y el impacto de los eventos financieros en el tiempo (Trna & Giménez-Martínez, 2012).

Por el otro lado, las DNN se utilizan para descubrir patrones complejos y relaciones no lineales en extensos conjuntos de datos financieros. Estas redes son eficaces para detectar señales ocultas en el "ruido" del mercado, realizar análisis de riesgo y optimizar la asignación de activos en una cartera. Al analizar una amplia gama de factores financieros y económicos, las DNN

pueden proporcionar una comprensión profunda y matizada de las dinámicas del mercado (Floratos et al., 2022).

El modelo de Dong, Huang, Ma y Qian (2021) destaca por integrar de manera efectiva estas tecnologías en la toma de decisiones de inversión. El estudio demuestra que el modelo de aprendizaje profundo por refuerzo supera a las estrategias de inversión convencionales en términos de rentabilidad, aprovechando tanto la representación de datos en el tiempo mediante RNN como el análisis de patrones complejos a través de DNN.

El estudio de **Lee, Kim, Yi y Kang (2020)** introduce un enfoque distinto a los mencionados anteriormente para la gestión de carteras mediante el sistema MAPS (Multi-Agent reinforcement learning-based Portfolio management System). MAPS utiliza un modelo de aprendizaje por refuerzo multiagente, donde cada agente funciona como un inversor independiente, diseñando su propia cartera. Esta estrategia de inversión se centra en la diversificación y la maximización de recompensas individuales. El sistema fue entrenado con datos históricos de precios de cierre diarios de aproximadamente 3,000 empresas estadounidenses durante un período de 18 años, proporcionando un amplio contexto de mercado para el aprendizaje y la adaptación del sistema.

MAPS demuestra la eficacia de un enfoque diversificado en la gestión de carteras. Al operar con múltiples agentes que actúan como inversores independientes, el sistema logra una diversificación significativa, lo que resulta en una reducción del riesgo total. Los resultados indican que MAPS supera consistentemente a todas las estrategias de referencia en términos de retorno y ratio de Sharpe. Esta superioridad en el rendimiento ajustado al riesgo subraya la ventaja de utilizar múltiples agentes en la gestión de carteras, donde la diversidad de enfoques de inversión contribuye a una mayor estabilidad y rentabilidad (Lee et al., 2020).

El estudio de **Gao et al. (2021)** aborda la aplicación del algoritmo Deep Deterministic Policy Gradient (DDPG) en la gestión de portafolios. Este estudio se centra en el entorno de inversión, caracterizado por mercados financieros dinámicos y a menudo impredecibles, y busca maximizar el retorno de la inversión mediante la adaptación continua de la composición del portafolio a las condiciones cambiantes del mercado.

En el modelo DDPG, el 'actor' se encarga de decidir la asignación de activos en el portafolio, basándose en el estado actual del mercado, mientras que el 'crítico' evalúa estas decisiones, proporcionando una estimación del valor o retorno esperado de la distribución del portafolio. La idoneidad del DDPG para la gestión de portafolios se debe a su habilidad para operar en espacios de acción continua, lo que permite ajustes precisos en la asignación de activos y facilita una gestión eficaz de la inversión. Además, el algoritmo equilibra la exploración de nuevas estrategias de asignación de portafolio con la explotación de estrategias conocidas, utilizando ruido aditivo en la política del actor para una exploración efectiva del espacio de decisión (Guha, 2021).

Un aspecto crítico del aprendizaje del DDPG como explica Aitor López Sánchez (2021) es su capacidad para aprender de experiencias pasadas y tendencias históricas del mercado a través de un buffer de repetición, lo cual es esencial en mercados financieros donde las tendencias históricas pueden ofrecer información valiosa. Por otro lado, las actualizaciones graduales de los parámetros de las redes ayudan a mantener la estabilidad en un entorno financiero que puede ser muy volátil, resaltando la importancia de la adaptabilidad y la precisión en la gestión de portafolios.

La gestión de carteras de inversión ha experimentado una transformación significativa con la incorporación del Deep Reinforcement Learning (DRL). Los estudios de Gao et al. (2021), Dong et al. (2021), Lee et al. (2020) y nuevamente Gao et al. (2021) ilustran un avance notable

en la aplicación de esta tecnología, ofreciendo soluciones innovadoras para enfrentar la dinámica y la incertidumbre de los mercados financieros.

Los avances en DRL, ilustrados por estos estudios, se centran en mejorar la capacidad de los modelos para analizar y reaccionar ante las condiciones cambiantes del mercado. Gao et al. (2021) introducen el Deep Q-Network (DQN) y el Deep Deterministic Policy Gradient (DDPG), destacando la importancia de estructuras jerárquicas y decisiones basadas en la valoración continua de las recompensas futuras.

Por otro lado, Dong et al. (2021) llevan esta innovación un paso más allá al combinar Redes Neuronales Recurrentes (RNN) y Redes Neuronales Profundas (DNN). Este enfoque permite una interpretación más profunda de los datos del mercado, aprovechando la capacidad de las RNN para analizar secuencias temporales y de las DNN para identificar patrones complejos, lo que resulta en una toma de decisiones más informada y estratégica.

La adaptabilidad es un componente crucial en la gestión de carteras, especialmente en mercados caracterizados por su alta volatilidad. Aquí, el sistema MAPS de Lee et al. (2020) demuestra su fortaleza, utilizando un enfoque de aprendizaje por refuerzo multiagente que permite a cada agente actuar como un inversor independiente, promoviendo la diversificación y la minimización del riesgo. Este enfoque multiagente mejora la adaptabilidad, además de aumentar el rendimiento ajustado al riesgo. Por último, la investigación de Gao et al. (2021) con su modelo DDPG también enfatizan la adaptabilidad, equilibrando la exploración de nuevas estrategias de asignación de cartera con la explotación de enfoques probados. Esto permite al modelo tener la capacidad para aprender de experiencias pasadas y adaptarse a condiciones cambiantes.

2.3. Deep Reinforcement Learning en Trading

El Trading, entendido como la actividad de comprar y vender activos financieros como acciones, bonos, divisas, entre otros, representa uno de los campos más desafiantes y dinámicos en el mundo financiero. La naturaleza impredecible de los mercados financieros, junto con la necesidad de tomar decisiones rápidas y eficientes, hace del trading un candidato ideal para la aplicación de tecnologías avanzadas como el Deep Reinforcement Learning (Lee et al., 2021).

El DRL, subcampo del aprendizaje automático, ha experimentado un crecimiento notable en su aplicación al trading, ofreciendo soluciones innovadoras y sofisticadas para la toma de decisiones en ambientes de alta incertidumbre y variabilidad (Guerra, 2023). Estudios recientes, como el de Bajpai (2021), han demostrado cómo el DRL puede ser eficaz en el desarrollo de estrategias de trading algorítmicas, adaptándose a las condiciones cambiantes del mercado y aprendiendo de la experiencia para mejorar el rendimiento de las operaciones.

La capacidad del DRL para manejar y procesar grandes cantidades de datos de mercado en tiempo real, junto con su habilidad para aprender de interacciones complejas y no lineales en los datos financieros, lo convierte en una herramienta poderosa para el trading (Guerra, 2023). Por ejemplo, investigaciones como la de Hirsa et al. (2021) han explorado el uso del DRL en la optimización de estrategias de trading intradía y la gestión de riesgos, demostrando mejoras significativas en la rentabilidad y reducción de pérdidas en comparación con métodos tradicionales (Hirsa et al., 2021).

Además, el DRL ofrece ventajas únicas en la adaptación a los mercados financieros (Gao et al., 2021). Estudios como el de Ma, Wang y Fleiss (2021) aplican técnicas avanzadas de DRL para predecir movimientos de precios y optimizar la ejecución de órdenes, destacando la

capacidad del DRL para aprender patrones de datos históricos, además de también ajustarse a las nuevas tendencias y anomalías del mercado.

En la siguiente sección, profundizaremos en los desarrollos más relevantes del DRL aplicado al trading. Examinaremos los estudios clave que han marcado el avance en este campo, analizando tanto las técnicas específicas de DRL empleadas como sus resultados prácticos. Este análisis proporcionará una base sólida para entender cómo el DRL está redefiniendo las estrategias de trading y contribuyendo al desarrollo de sistemas de trading automatizados más inteligentes y eficientes.

En el ámbito del trading financiero, el estudio de **Tsantekidis, Passalis y Tefas (2020)** introduce una metodología vanguardista que emplea la destilación de redes neuronales dentro del Deep Reinforcement Learning (DRL) para mejorar significativamente la eficacia y el rendimiento de los agentes de trading. Esta técnica facilita la transmisión de conocimientos desde modelos complejos y de gran tamaño hacia otros más compactos y operativamente eficientes, optimizando el consumo de recursos computacionales sin comprometer la calidad de las decisiones de trading. La investigación destaca por su contribución al desarrollo de estrategias de inversión automatizadas, capaces de adaptarse dinámicamente a las fluctuantes condiciones del mercado financiero, ofreciendo así una notable mejora en la precisión y adaptabilidad de los sistemas de trading algorítmico. La relevancia de este trabajo se manifiesta en su potencial para revolucionar las prácticas de trading, mediante la implementación de soluciones tecnológicas avanzadas que responden eficientemente a los retos presentados por la volatilidad y la complejidad inherentes a los mercados financieros.

El estudio conducido por **Lee, Koh y Choe (2021)** representa un hito significativo en la optimización de carteras y sistemas de trading a través del uso del Deep Reinforcement Learning, subrayando la influencia crítica de la composición de los datos de entrada, la configuración de la red de aprendizaje, y el establecimiento de recompensas adecuadas. La investigación propone una estructura de aprendizaje profundo que inicialmente emplea la transformación de ondículas (Wavelet Transformation, WT) para depurar los datos temporales de precios de acciones de ruido, utilizando posteriormente únicamente los datos de la onda madre (alta frecuencia) como entrada. Este proceso es seguido por el aprendizaje reforzado utilizando los mencionados datos de alta frecuencia, con una red que emplea Long Short-Term Memory (LSTM) para determinar acciones, que pueden ser decididas por la red LSTM o generadas aleatoriamente. Además, se aprende el sistema óptimo de trading de inversión a través de las acciones de una transacción dada y recompensas apropiadas, mejorando el desempeño del trading sin necesidad de construir un modelo predictivo. La validación de este enfoque se llevó a cabo mediante la aplicación de índices como el S&P500, DJI y KOSPI200, demostrando mejoras significativas en el rendimiento del trading, particularmente en mercados altamente volátiles, y resaltando la importancia de una composición adecuada de los datos de entrada, configuraciones de la red de aprendizaje y la definición de recompensas en el DRL.

En el artículo "Deep Reinforcement Learning Pairs Trading with a Double Deep Q-Network", **Brim (2020)** explora el potencial del aprendizaje profundo por refuerzo (DRL) para la optimización de estrategias de trading de pares, implementando una innovadora arquitectura conocida como Double Deep Q-Network (DDQN). Este estudio se sitúa en la vanguardia de la investigación financiera aplicada, al integrar avanzadas técnicas de aprendizaje automático para mejorar la toma de decisiones en el trading de pares, una estrategia que busca capitalizar las discrepancias de precios entre dos activos financieramente correlacionados.

La investigación aborda los desafíos inherentes al RL y al trading de pares mediante la incorporación de DDQN, una mejora significativa sobre el modelo Deep Q-Network (DQN) tradicional. El DDQN corrige la tendencia del DQN a sobreestimar los valores Q, a través de la

utilización de dos redes neuronales que separan la selección de la mejor acción de la evaluación de esa acción, proporcionando así estimaciones de valor más precisas y fomentando una política de trading más estable y confiable.

Brim (2020) detalla cómo la transformación de ondículas se aplica para preprocesar los datos de entrada, una técnica esencial para filtrar el ruido y extraer características significativas de las series temporales de precios de acciones. Esta preparación de los datos, combinada con el enfoque de DDQN, permite al agente de DRL identificar oportunidades de trading de pares con mayor precisión, maximizando la recompensa acumulada mediante la adaptación dinámica a las condiciones cambiantes del mercado.

El artículo contribuye significativamente a la literatura existente sobre DRL y trading financiero, proporcionando evidencia empírica de la eficacia del modelo DDQN en el contexto del trading de pares. Los experimentos realizados por Brim, utilizando índices como el S&P 500, demuestran la superioridad del enfoque DDQN sobre estrategias convencionales y modelos DQN estándar, en términos de rendimiento financiero y gestión de riesgos.

Otro estudio pionero en este campo es el realizado por **Hao, L., Wang, B., Lu, Z., & Hu, K. (2022)**, quienes en su artículo "Application of Deep Reinforcement Learning in Financial Quantitative Trading", exploran la viabilidad de aplicar el aprendizaje profundo por refuerzo (Deep Reinforcement Learning, DRL) para mejorar la toma de decisiones en el trading cuantitativo. Inspirándose en la teoría de juegos de decisión que catapultó a AlphaGo al éxito, los autores proponen la utilización del algoritmo Deep Q Network (DQN) para capturar las complejas dinámicas del mercado financiero y, en consecuencia, maximizar los beneficios mediante la adopción de estrategias de trading más precisas y efectivas.

El trabajo de Hao et al. (2022) se centra en la implementación del algoritmo DQN, resaltando su capacidad para manejar efectivamente problemas de aprendizaje con objetivos a largo plazo y recompensas diferidas. Un aspecto crucial de su estudio es el diseño meticuloso de la función de recompensa, que permite al modelo identificar dependencias ocultas y dinámicas en los datos del mercado de valores, una contribución significativa al campo del trading cuantitativo.

Además, los autores adoptan una estrategia de trading basada en el promedio doble, utilizando medias móviles de 5 y 15 días para determinar los puntos óptimos de compra y venta. Esta estrategia no solo mejora la precisión de la toma de decisiones, sino que además subraya la importancia de integrar indicadores técnicos dentro del marco del aprendizaje por refuerzo, para dirigir las acciones del agente de manera más efectiva. La comparación del algoritmo DQN con variantes como Double-DQN y Dueling-DQN destaca la superioridad del primero en el contexto del trading cuantitativo, tanto en términos de retorno de inversión como de beneficios generados. Este hallazgo valida la aplicabilidad del DQN en el ambiente de toma de decisiones de trading, demostrando su potencial para revolucionar el trading cuantitativo mediante la maximización de beneficios.

El estudio de Hao et al. (2022) contribuye significativamente al estado del arte la integración de estrategias de trading basadas en indicadores técnicos en el trading cuantitativo, demostrando la efectividad del aprendizaje profundo por refuerzo, en particular el algoritmo DQN, para capturar las dinámicas complejas del mercado y mejorar la toma de decisiones de trading.

A través de la revisión de estos estudios recientes, se observa una diversidad de enfoques y aplicaciones del DRL, reflejando su versatilidad y potencial en la mejora del rendimiento de trading. La metodología de destilación de redes neuronales propuesta por Tsantekidis, Passalis y Tefas (2020) subraya la importancia de la eficiencia computacional, permitiendo la transmisión de conocimientos desde modelos complejos a otros más compactos, optimizando así el consumo

de recursos sin comprometer la calidad de las decisiones de trading. Este enfoque contrasta con el uso del algoritmo Deep Q Network (DQN) por Hao et al. (2022), que destaca por su capacidad para manejar problemas de aprendizaje con objetivos a largo plazo, integrando indicadores técnicos para mejorar la precisión en la toma de decisiones.

Por otro lado, el estudio de Lee, Koh y Choe (2021) introduce una estructura de aprendizaje que combina la transformación de ondículas con datos de alta frecuencia, utilizando la memoria a largo y corto plazo (LSTM) para determinar acciones, lo que mejora el desempeño del trading especialmente en mercados volátiles. Esta optimización de carteras mediante una adecuada preparación de datos y configuraciones de red específicas ofrece un contraste interesante con la investigación de Brim (2020), que se enfoca en el trading de pares utilizando la arquitectura Double Deep Q-Network (DDQN) para corregir la sobreestimación de los valores Q y mejorar la estabilidad de la política de trading.

La comparación de estas metodologías revela un tema común: la importancia de una preparación y análisis de datos meticulosos, así como la configuración de la red y la función de recompensa, en la eficacia del DRL para el trading financiero. Estos estudios colectivamente demuestran la potencialidad del DRL en redefinir las estrategias de trading y contribuir al desarrollo de sistemas de trading automatizados más inteligentes y eficientes. Aunque cada enfoque tiene sus ventajas y limitaciones, juntos ofrecen una vista comprensiva sobre cómo el DRL puede ser aplicado de manera efectiva en el trading financiero.

Capítulo 3

Alcance de la tesis

Este proyecto propone un análisis comparativo entre la configuración predeterminada de la tasa de aprendizaje en algoritmos de Deep Reinforcement Learning, específicamente en el marco de Stable Baselines 3 utilizado para la gestión de portafolios financieros, y una versión modificada de este algoritmo con una tasa de aprendizaje ajustada al alza. El objetivo es optimizar la gestión de carteras de inversión en entornos financieros dinámicos.

Se evaluarán ambas versiones del algoritmo en términos de rendimiento y riesgo, con un enfoque particular en el ratio de Sharpe como medida de rendimiento ajustado al riesgo. La evaluación busca determinar si una tasa de aprendizaje elevada puede conducir a una mejora significativa en la gestión de portafolios financieros, superando así los benchmarks establecidos por la configuración predeterminada en Stable Baselines 3.

3.1. Hipótesis

h.1 Hipótesis Principal: La hipótesis técnica central de esta investigación sostiene que una versión del algoritmo de Deep Reinforcement Learning con una tasa de aprendizaje ajustada al alza superará significativamente en rendimiento, medido por el ratio de Sharpe, a la configuración predeterminada en Stable Baselines 3. Esto se basa en la premisa de que una adaptación más ágil a las condiciones cambiantes del mercado puede ser crucial para una gestión de portafolios más efectiva.

La hipótesis técnica principal, denotada como H1, se define matemáticamente de la siguiente manera:

Definimos RS(A, α) como el ratio de Sharpe alcanzado por el algoritmo A de Deep Reinforcement Learning con una tasa de aprendizaje α. Sea α_0 la tasa de aprendizaje predeterminada en Stable Baselines 3 y $\alpha_1 > \alpha_0$ una tasa de aprendizaje ajustada al alza. Entonces, la hipótesis principal H1 puede expresarse matemáticamente como:

$$RS(A, \alpha_1) > RS(A, \alpha_0)$$

Ecuación 1. Hipótesis principal

3.2. Objetivos

o.1 Desarrollar y aplicar mejoras específicas al algoritmo de aprendizaje reforzado profundo en FinRL para optimizar la selección y gestión de una cartera de activos financieros.
o.2 Realizar una comparación exhaustiva del rendimiento y el riesgo entre la cartera gestionada por los algoritmos con distintos learning rates.
o.3 Examinar la efectividad de las mejoras introducidas en términos de rentabilidad ajustada por el riesgo.
o.4 Documentar de manera detallada la metodología, implementación y resultados del modelo.

3.3. Asunciones

a.1 Disponibilidad de datos históricos y en tiempo real de precios de acciones y otros activos financieros.
a.2 Continuidad en la dinámica del mercado financiero que permite la aplicabilidad de aprendizajes pasados.
a.3 Los recursos computacionales actuales son suficientes para procesar y analizar los datos requeridos.
a.4 La eficacia del aprendizaje reforzado profundo se mantiene a pesar de las variaciones y volatilidades del mercado, permitiendo que el algoritmo mejorado se adapte y responda eficientemente a condiciones cambiantes.

3.4. Restricciones

r.1 Límite de tiempo: Completar el proyecto dentro de 100 horas laborales.
r.2 Sin financiamiento adicional para recursos o software especializado.
r.3 Dependencia de las capacidades computacionales del equipo disponible, sin acceso a hardware de alto rendimiento.
r.4 Acceso limitado a conjuntos de datos completos y actualizados en tiempo real, esencial para la precisión del aprendizaje reforzado profundo, lo cual puede limitar la capacidad de evaluación y ajuste fino del algoritmo mejorado.
r.5 La complejidad del desarrollo e integración de mejoras al algoritmo existente en FinRL, dadas las limitaciones de tiempo y recursos computacionales, puede restringir el alcance de las optimizaciones realizables.

Capítulo 4

Marco teórico

El marco teórico de esta investigación se adentra en la exploración de conceptos y teorías fundamentales que forman la base del estudio. Comenzando con la Teoría de Gestión de Carteras de Markowitz, muestra la percepción de la inversión a través de la diversificación y la correlación entre activos, esta sección establece las bases del portfolio management. Posteriormente, se profundiza en el ámbito del Aprendizaje Profundo y el Aprendizaje por Refuerzo, destacando su significativa influencia en el desarrollo de tecnologías de inteligencia artificial avanzadas. La convergencia de estos campos en el Deep Reinforcement Learning es examinada, resaltando su capacidad para abordar complejas problemáticas de toma de decisiones y optimización. Además, se discuten aspectos críticos como la tasa de aprendizaje y el algoritmo de Proximal Policy Optimization (PPO), ilustrando su relevancia en la mejora y eficacia de los modelos de DRL.

4.1. Teoría de Gestión de carteras de Markowitz

La teoría sobre la gestión de carteras, formulada por Harry Markowitz en 1952, representó una revolución en el ámbito financiero (Pedram Nezafat, 2023). Anterior a este desarrollo, las inversiones se analizaban desde una perspectiva individual por activo. Con la introducción de esta teoría, Markowitz destacó la relevancia de diversificar y entender la correlación entre los activos dentro de una cartera. Esta innovación modificó el método utilizado por inversores y administradores de carteras para evaluar los activos financieros y sentó las bases para futuros avances y estrategias en la gestión de inversiones. Previamente a la teoría de Markowitz, la estrategia de inversión se concentraba en el análisis independiente de cada activo, donde se valoraban elementos como el desempeño histórico, la solidez financiera de la empresa y las proyecciones del sector, sin considerar la interacción entre distintas inversiones en una misma cartera. Según Jose (2017), este método no proporcionaba una perspectiva completa sobre el potencial y los riesgos de una cartera.

Harry Markowitz introdujo dos conceptos clave en la gestión de carteras: la diversificación y la correlación entre los activos. Antes de su aporte, el enfoque de inversión se centraba en la rentabilidad individual de los activos. Sin embargo, Markowitz propuso que al combinar activos con correlaciones bajas o negativas se podría disminuir el riesgo de la cartera sin comprometer el rendimiento esperado. Su teoría subrayó la importancia de evaluar cómo los activos interactúan entre sí bajo diversas condiciones del mercado. Además, presentó un modelo cuantitativo para la

selección de carteras que equilibra el riesgo y el retorno, descrito en su publicación de 1952, el cual permitía a los inversores calcular la combinación óptima de activos considerando tanto el rendimiento esperado como la volatilidad (Markowitz, 1952).

4.2. Deep Reinforcement learning

El Deep Reinforcement Learning, es una combinación de dos campos poderosos en machine learning: el aprendizaje profundo (Deep Learning) y el aprendizaje por refuerzo (Reinforcement Learning). Para entenderlo bien, vamos a desglosar ambos componentes y luego explorar cómo se unen en el Deep Reinforcement Learning.

El **aprendizaje profundo** es una subcategoría del machine learning que utiliza redes neuronales con muchas capas (de ahí el "profundo" en su nombre). Estas redes son estructuras computacionales inspiradas en el funcionamiento del cerebro humano, diseñadas para reconocer patrones complejos en grandes cantidades de datos (IBM, s.f). A diferencia de las metodologías tradicionales de machine learning, que requieren datos estructurados y la definición manual de características, el Deep Learning automatiza la extracción de características relevantes, simplificando considerablemente la preparación de los datos y mejorando la eficiencia en la resolución de tareas complejas (IBM, s.f.).

Las redes neuronales profundas aprenden de manera supervisada, lo que significa que necesitan un conjunto de datos etiquetados para entrenarse (IBM, s.f.). A través del entrenamiento, ajustan sus parámetros internos para minimizar el error en sus predicciones o clasificaciones, mejorando así su precisión en tareas específicas. Los modelos de Deep Learning se autoajustan y perfeccionan a través de técnicas avanzadas como el Back-Propagation (BP) y el descenso de gradiente, permitiendo un aprendizaje profundo y adaptativo sin necesidad de intervención humana en la definición de jerarquías de características. Esta característica les otorga una precisión superior en tareas de clasificación y reconocimiento complejas, marcando una diferencia notable respecto a los enfoques más tradicionales (IBM, s.f.).

Las redes neuronales se fundamentan en una estructura compuesta principalmente por un conjunto específico de unidades de procesamiento, denominadas neuronas. Cada neurona (véase Ilustración 1 como referencia) actúa como un elemento de procesamiento esencial, encargada de recibir entradas que son sumadas y ponderadas. Este valor acumulado es luego transformado mediante una función de activación, resultando en una salida específica. Esta salida puede, a su vez, ser dirigida hacia otras neuronas a través de enlaces ponderados conocidos como sinapsis, facilitando así la interconexión entre las distintas unidades de procesamiento (Brunton et al., 2019).

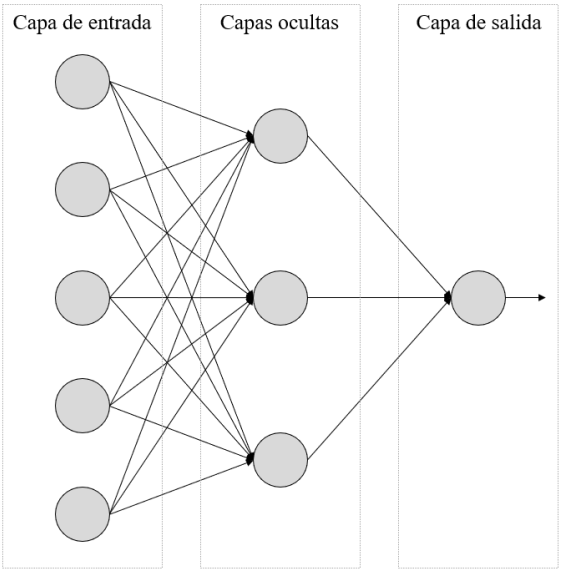


Ilustración 1. Esquema ilustrativo de una red neuronal de una capa oculta

Fuente: Elaboración propia

El diseño convencional de una red neuronal se estructura en torno a tres capas diferenciadas, cada una integrada por diversas neuronas que cumplen funciones específicas. En la capa de entrada, se introducen los datos al sistema. Seguidamente, la capa oculta se encarga de procesar estos datos, realizando para cada neurona una operación de suma ponderada, seguida de la aplicación de una función de activación seleccionada para adecuarse a los requerimientos del proceso. Finalmente, la capa de salida genera la respuesta final de la red, culminando así el proceso de cálculo (Xiao et al., 2020). Desde una perspectiva matemática, las redes neuronales se representan como:

$$y = f(x) = \sum_{j=1}^{J} w_j \, \delta \left(\sum_{i=1}^{I} w_{ij} \cdot x_i \cdot b_j \right) + \beta + \varepsilon$$

Ecuación 2. Representación matemática de una red neuronal

Donde:

- w_{ij} *es el factor de peso entre la neurona i (entrada) con la neurona j (oculta)*
- w_j *es el factor de peso entre la neurona j (oculta) con la de salida*
- δ *es la función de activación*
- β *es el sesgo de para la neurona de salida*
- ε *es error aleatorio*

Como nos explica Ti et al. (2020), las redes neuronales se clasifican dentro del espectro de métodos pertenecientes al aprendizaje supervisado, siendo su entrenamiento ejecutado mediante el algoritmo de Back-Propagation (BP). Este algoritmo desempeña un papel crucial en la optimización de los pesos sinápticos de la red, con el objetivo primordial de reducir el error asociado a las predicciones de la red neuronal. El proceso se estructura en dos fases distintas: la propagación hacia adelante (forward propagation), en la cual la señal se transmite desde la capa

de entrada a través de las capas ocultas hasta la capa de salida, y la retro-propagación del error (back-propagation), que ajusta los pesos en función del error calculado.

Durante el entrenamiento, las variables de entrada, al ser ponderadas por sus respectivos pesos, avanzan a través de las capas ocultas hacia la capa de salida en la fase de propagación hacia adelante. La discrepancia entre los valores observados de la variable de respuesta y las predicciones de la red se evalúa mediante una función de pérdida, la cual mide la precisión con la que la red está alcanzando su objetivo. Posteriormente, en la fase de retro-propagación, se ajustan los pesos sinápticos de la red con el fin de acercar las salidas generadas a los valores reales deseados. Este proceso de ajuste y evaluación se repite iterativamente hasta que el modelo alcanza un nivel de predicción que se considera satisfactorio. Esta metodología permite a las redes neuronales aprender de manera efectiva, ajustándose a los patrones subyacentes en los datos de entrenamiento, lo que las habilita para realizar predicciones o clasificaciones precisas sobre nuevos conjuntos de datos (Ti et al. 2020). La ecuación que refleja la salida de una red neuronal del tipo back-propagation es la siguiente:

$$(\hat{y}_i)_k = \sigma^{out}\left(\sum_{i=1}^{n_H} w_j^{out} \cdot x_i^H \cdot b_j^{out}\right), \quad k = 1,2,,\dots,l,$$

Ecuación 3. Salida de una red neuronal del tipo back-propagation

Donde:

- σ^{out} *es la función de activación de la capa de salida*
- w_j^{out} *son los pesos de la capa de salida*
- b_j^{out} *son los sesgos de la capa de salida*

El algoritmo de Back-Propagation se fundamenta en el principio de avanzar en dirección opuesta al gradiente de la función objetivo, un enfoque conocido como método de descenso de gradiente. Inicia desde una posición arbitraria, aspirando a alcanzar el punto más bajo de la función o aquel en el que el gradiente se reduce a cero. Este procedimiento permite la actualización secuencial de los parámetros de entrenamiento, específicamente los pesos (w) y los sesgos (b), a través de las distintas capas de la red, empleando el método de descenso de gradiente. La esencia de este proceso radica en la propagación del gradiente de error en sentido inverso, comenzando desde la capa de salida y avanzando hacia la capa de entrada (Dean, 2014).

La implementación del método de descenso de gradiente implica la necesidad de determinar la magnitud de cada paso que se da en dirección opuesta al gradiente, conocida como tasa de aprendizaje o learning rate. Una estrategia común consiste en ajustar este avance en función de la magnitud del gradiente, permitiendo así un ajuste más fino de los parámetros a medida que el algoritmo itera hacia la optimización de la red neuronal (Dean, 2014). La fórmula para esta metodología de ajuste gradual es la siguiente:

$$\mathrm{w}_{n+1} = \mathrm{w}_n - \lambda \nabla f(\mathrm{w}_n)$$

Ecuación 4. Ajuste gradual del gradiente

Donde:

- w_{n+1} *es el vector de pesos en el momento n+1*
- w_n *es el vector de pesos en el momento n (actual)*
- λ *es la tasa de aprendizaje*
- ∇f *es el gradiente de la función objetivo*

Por otro lado, **el aprendizaje por refuerzo** se caracteriza por su capacidad para abordar problemáticas asociadas a la toma de decisiones secuenciales. Huang, Chang, y Chakraborty (2019) nos explican que este paradigma se sustenta en la premisa de que un agente decisional interactúa dinámicamente con su entorno, proceso ilustrado en la Ilustración 2. Dicho agente, encargado de ejecutar acciones, recibe retroalimentación de su entorno en forma de recompensas o penalizaciones, lo cual le permite ajustar sus estrategias de decisión de manera progresiva para optimizar un objetivo predeterminado.

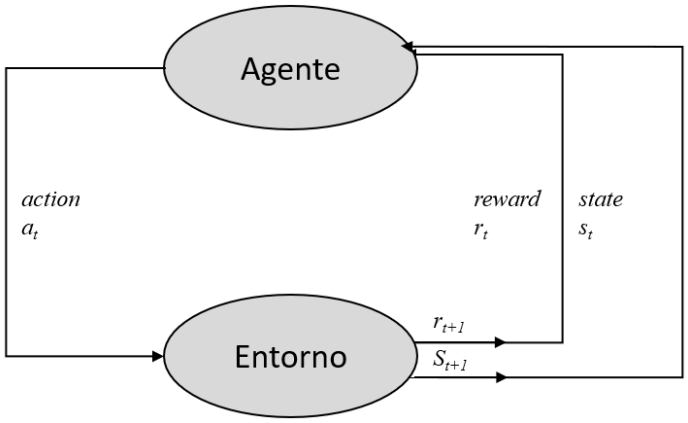


Ilustración 2. Esquema ilustrativo de la relación entre el agente y su entorno

Fuente: Elaboración propia

Las interacciones entre el agente y el entorno se desarrollan de manera secuencial a lo largo del tiempo, denotado como t = 0, 1, 2, …, t. En cada momento específico dentro de esta secuencia temporal, el agente adquiere una representación del estado actual del entorno, simbolizado como S_t, sobre la base del cual selecciona una acción, A_t, para ejecutar. La implementación de esta acción conlleva la transición del entorno hacia un nuevo estado, denotado como S_{t+1}, y resulta en la obtención de una recompensa por parte del agente, R_{t+1}, que refleja la eficacia de la acción previamente seleccionada. Este mecanismo de recompensa provee al agente una indicación sobre la idoneidad de las acciones emprendidas, facilitando un aprendizaje fundamentado en el principio de ensayo y error. A través de este proceso iterativo, el agente afina sus estrategias decisionales para optimizar la acumulación de recompensas futuras (García, 2020).

Los Procesos de Decisión de Markov (MDP) se establecen como una metodología clave para la estructuración de la toma de decisiones en secuencias. Un MDP se define mediante una quíntuple (S, A, P, R, γ), donde S designa un conjunto de estados disponibles. En un estado dado, se selecciona una acción, A_t, de un conjunto total de acciones A. La implementación de esta acción

induce la transición del sistema desde el estado actual, S_t, al siguiente estado, S_{t+1}, con una determinada probabilidad de transición, $P(S_{t+1}/A_t, S_t)$. Dicha acción acarrea una recompensa, R_{t+1}. El componente γ, que se halla en el rango [0,1], regula el peso de las recompensas presentes frente a las futuras, permitiendo una valoración ajustada de las decisiones en el marco de sus repercusiones a largo plazo (Pröllochs & Feuerriegel, 2018).

El Deep Reinforcement Learning como mencionado en la introducción de esta sección, representa un punto de convergencia entre dos de las áreas más dinámicas y potentes de machine learning: el aprendizaje profundo (Deep Learning, DL) y el aprendizaje por refuerzo (Reinforcement Learning, RL). Esta sinergia capitaliza las fortalezas de cada campo haciendo posible ahora abordar problemas complejos de decisión y optimización que eran inaccesibles con enfoques anteriores. En esencia, DRL emplea la capacidad de las redes neuronales profundas para interpretar estados ambientales complejos, junto con la estructura de decisión y aprendizaje basada en recompensas del aprendizaje por refuerzo, creando sistemas capaces de aprender y adaptarse a entornos dinámicos (Garrido, 2024).

El DL proporciona a los modelos de DRL una herramienta para la extracción automática de características y el reconocimiento de patrones en datos de alta dimensión. Este aspecto es crucial para el procesamiento y la interpretación de estados ambientales complejos, permitiendo que los agentes de RL comprendan y naveguen por su entorno (Dean. 2014).

Por otro lado, el RL aporta el marco conceptual y algorítmico para que los agentes aprendan a tomar decisiones óptimas a través de la interacción con su entorno. La capacidad de los agentes de RL para aprender de las consecuencias de sus acciones y ajustar sus estrategias de comportamiento en consecuencia es aumentada gracias a la capacidad de procesamiento de información de las redes neuronales profundas. Este aprendizaje basado en la interacción posibilita que los sistemas de DRL se adapten y mejoren continuamente su rendimiento en tareas complejas, como la navegación autónoma, juegos estratégicos, la optimización de procesos y el trading en mercados financieros (García, 2020).

4.3. Learning Rate

La tasa de aprendizaje es un hiperparámetro crítico para el DRL, que juega un papel fundamental en la convergencia de los algoritmos hacia una solución óptima. Este parámetro indica el tamaño del paso que el algoritmo de descenso de gradiente toma hacia el óptimo local, influenciando directamente la eficiencia y la eficacia del proceso de entrenamiento de los modelos de redes neuronales (Zvornicanin & Zvornicanin, 2024).

En esencia, la tasa de aprendizaje determina la rapidez con la que un algoritmo ajusta sus parámetros en respuesta a la estimación del error cada vez que los parámetros son actualizados. Un valor óptimo para la tasa de aprendizaje es crucial, ya que valores demasiado bajos pueden resultar en una convergencia excesivamente lenta hacia el óptimo, mientras que valores excesivamente altos pueden causar divergencia, impidiendo que el algoritmo alcance una solución viable (Zvornicanin & Zvornicanin, 2024). A continuación, en la Ilustración 3 se muestra esta comparativa gráficamente:

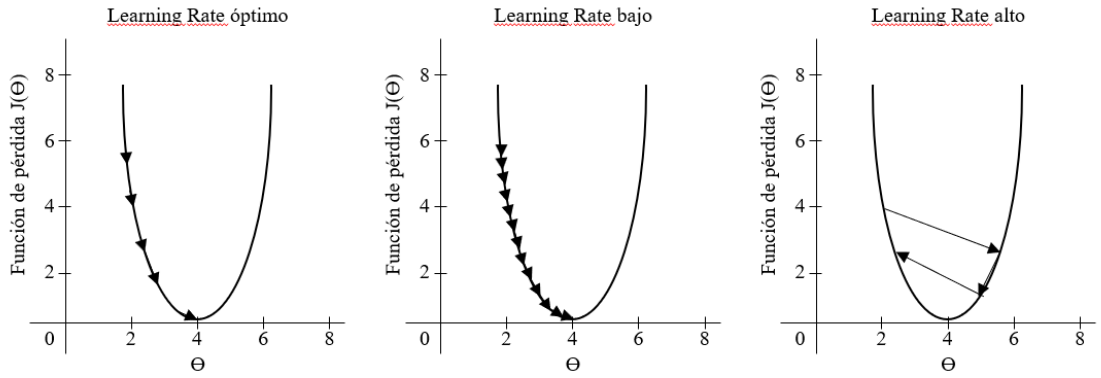


Ilustración 3. Representación gráfica del efecto del Learning Rate en la minimización de la función de pérdida

Fuente: Elaboración propia

El proceso de fine-tuning de este hiperparámetro es esencial, especialmente en conjuntos de datos de gran escala, donde una elección inadecuada puede llevar a resultados subóptimos o a la falta de convergencia. Existen métodos adaptativos para la optimización de la tasa de aprendizaje, tales como el Annealing, el cual es uno de los enfoques más simples que podemos implementar que consiste en reducir la tasa de aprendizaje a medida que avanzan los epoch (número de veces que se van a pasar cada ejemplo de entrenamiento por la red). Otra alternativa es RMSProp, el cual es un optimizador que actualizará la tasa de aprendizaje durante las iteraciones, basándose en el hecho de que las actualizaciones tempranas deberían tener menos peso en las actualizaciones posteriores. Además de estas metodologías existen muchas otras como Adam, Adagrad, Adadelta o PBT entre otros, los cuales ajustan dinámicamente el learning rate durante el entrenamiento para mejorar la convergencia (Barreto & Barreto, 2024). Estos métodos buscan equilibrar la exploración del espacio de parámetros con la eficiencia en la convergencia hacia mínimos de la función de pérdida.

4.4. Proximal Policy Optimization

El algoritmo de Proximal Policy Optimization (PPO) representa un avance significativo en el DRL, ofreciendo una solución eficaz para la optimización de políticas en entornos de control continuo y en la gestión de espacios de acción amplios. Desarrollado por investigadores de OpenAI en 2017, el PPO ha emergido como uno de los algoritmos más populares y efectivos en DRL, destacándose por su balance entre simplicidad de implementación y robustez en el rendimiento (González Oviedo, 2023).

El diseño del PPO se basa en mejorar y simplificar los métodos tradicionales de gradiente de política. A diferencia de su predecesor, el Trust Region Policy Optimization (TRPO), que requiere cálculos complejos para asegurar actualizaciones seguras de la política, PPO introduce un enfoque más accesible sin comprometer la eficacia. La clave de este enfoque radica en la función de pérdida "recortada" y una función objetivo que limita las actualizaciones de la política a un rango aceptable, previniendo cambios abruptos y potencialmente perjudiciales para el aprendizaje (Schulman et al., 2017).

Como nos explica Ignacio Such Ballester (2024) se debe tomar en cuenta la ratio de probabilidades que existe entre las políticas antiguas y las recientes, que se expresa mediante:

$$rt(\theta) = \frac{\pi_\theta(at\,|st)}{\pi\theta_{Old}(at\,|st)}$$

Ecuación 5. Ratio de probabilidades. Políticas antiguas y recientes

Donde:

- $\pi_\theta(at\,|st)$ *simboliza la política implementada por la red neuronal.*

Esta ratio, también conocida como muestreo de importancia, tiene el potencial de limitar las actualizaciones excesivamente grandes. Shulman et al. (2017) introduce una función clip que limita el tamaño de las actualizaciones para prevenir la sobrevaloración de una acción. Esto conduce a la fórmula de PPO de la siguiente manera:

$$L^{CLIP}(\theta) = \hat{E}t[min(r_t(\theta))\hat{A}_t, clip(r_t(\theta),1-\varepsilon,1+\varepsilon)\hat{A}_t)]$$

Ecuación 6. Representación matemática del PPO

Donde:

- θ *es el parámetro de política*
- $\hat{E}t$ *denota la expectativa empírica sobre pasos temporales.*
- r_t *es la razón de la probabilidad bajo las nuevas y viejas políticas, respectivamente.*
- $\hat{A}_t$ *es la ventaja estimada en el tiempo.*
- ε *es un hiperparámetro, usualmente* 0.1 *o* 0.2.

Este enfoque implementa una manera de realizar una actualización de Región de Confianza compatible con el Descenso de Gradiente Estocástico, simplificando el algoritmo al eliminar la penalización KL y la necesidad de realizar actualizaciones adaptativas. En pruebas, este algoritmo ha mostrado el mejor rendimiento en tareas de control continuo y casi iguala el rendimiento de ACER (Actor-Critic with Experience Replay) en Atari, a pesar de ser mucho más simple de implementar (Schulman et al., 2017).

La versatilidad de PPO se ha demostrado en una variedad de aplicaciones, incluyendo el control de vehículos autónomos y la optimización de sistemas de fabricación, demostrando su capacidad para manejar desafíos complejos de control y optimización en diferentes dominios (Meyer, Robinson, Rasheed, & San, 2019; Zhu, Wang, & Zhang, 2020). Una de las ventajas principales de PPO sobre otros algoritmos de DRL es su balance entre rendimiento y simplicidad computacional, ofreciendo una implementación eficiente sin sacrificar la calidad del aprendizaje (Schulman et al., 2017). Además, según Such (2024), tiene aplicabilidad tanto en entornos de acción discretos como continuos.

Capítulo 5

Experimentos

5.1. Objetivo

El objetivo de esta investigación es explorar la influencia de la tasa de aprendizaje en la eficacia de algoritmos de Deep Reinforcement Learning aplicados a la gestión de portafolios financieros. Pretendemos determinar si una tasa de aprendizaje ajustada por encima del valor predeterminado en Stable Baselines 3 conduce a una mejora en el rendimiento de la gestión de portafolios, partiendo de la base de que una adaptación más ágil podría ser ventajosa en el dinámico entorno financiero.

5.2. Hipótesis

Hipótesis Nula (H0): No hay diferencia significativa o la tasa de aprendizaje predeterminada es igual o más efectiva que una tasa de aprendizaje elevada, medida por el ratio de Sharpe.

Hipótesis Alternativa (H1): Una tasa de aprendizaje elevada supera significativamente en rendimiento, medido por el ratio de Sharpe, a la tasa de aprendizaje predeterminada.

5.3. Implementación

La implementación de este Trabajo de Fin de Grado se ha llevado a cabo mediante la utilización de una serie de herramientas y librerías de Python para satisfacer las necesidades específicas del proyecto.

Entre las librerías de Python empleadas, *pandas* destaca por su potencia y flexibilidad en el manejo de datos, siendo utilizada ampliamente para la entrada y salida de archivos. Esta librería facilita una manipulación eficiente de conjuntos de datos tabulares a través de sus estructuras DataFrame y Series, lo cual ha sido esencial para la carga, limpieza y transformación de los datos utilizados en el proyecto. *Numpy*, por otro lado, es fundamental para la manipulación numérica, especialmente para operaciones con vectores y matrices. Su capacidad para realizar operaciones

matemáticas complejas ha sido clave en el manejo de grandes volúmenes de datos y la implementación de cálculos numéricos.

La visualización de resultados se ha realizado a través de *Matplotlib*, una librería que ha permitido la generación de gráficos, como los box plots, indispensables para la interpretación de los resultados del modelo. Por otra parte, la librería *DateTime* ha suministrado herramientas necesarias para la manipulación de fechas, un aspecto crucial en el manejo de series temporales financieras, facilitando el preprocesamiento de datos y la sincronización de series temporales.

Además de estas herramientas básicas, la implementación ha incorporado tecnologías especializadas en áreas de aprendizaje reforzado. *FinRL* ha sido utilizada para facilitar la implementación de estrategias de trading y la evaluación de modelos en contextos financieros reales, mientras que *StableBaselines3* ha proporcionado un marco robusto para la implementación de algoritmos de Deep Reinforcement Learning, permitiendo una experimentación eficaz y comparativa entre los distintos learning rates. Finalmente, *Gym* de OpenAI ha sido fundamental para el desarrollo del entorno de entrenamiento, ofreciendo la posibilidad de simular escenarios específicos en los que los modelos pueden ser entrenados y evaluados.

5.4. Diseño Experimental

Este trabajo investiga el impacto de la tasa de aprendizaje en el rendimiento de algoritmos de Deep Reinforcement Learning, específicamente el Proximal Policy Optimization (PPO) y Soft Actor-Critic (SAC), enfocándose en la optimización de carteras financieras. Se propone un experimento para determinar si las tasas de aprendizaje aplicadas a problemas financieros difieren en eficacia de aquellas utilizadas en contextos de robótica y videojuegos. Esta indagación se basa en la premisa de que una tasa de aprendizaje superior podría resultar más beneficiosa en el ámbito financiero debido a la naturaleza particular de los datos financieros.

Inicialmente, en el desarrollo de este proyecto, la **extracción de datos** financieros de alta calidad representa una etapa crítica. Estos datos son indispensables para entrenar nuestro modelo, proporcionando información relevante y actualizada que fundamenta el análisis y la toma de decisiones. Para la recolección de estos datos, hemos optado por utilizar YahooDownloader, siendo entonces Yahoo Finance el proveedor de datos, que nos ha facilitado el acceso a series temporales financieras completas de los activos incluidos en el índice DOW Jones 30. La descarga de datos se ha centrado en la obtención de datos *OHLCV* (Open, High, Low, Close, Volume) correspondientes a cada una de las empresas que forman parte del índice, abarcando un período desde el 1 de enero de 2008 hasta el 31 de diciembre de 2023. Esta extensa colección de datos constituye un amplio data frame para el análisis y posterior modelado. Conscientes de que los datos en bruto pueden no proporcionar toda la expresividad o información necesaria para un agente de aprendizaje reforzado, hemos procedido a enriquecer este conjunto de datos. Este enriquecimiento implica la adición de indicadores técnicos y una matriz de correlación, elementos que dotan al agente de información adicional crucial para la estimación de políticas de inversión efectivas manteniendo una gestión del riesgo eficaz.

A continuación, se **construye y diseña un entorno de entrenamiento**. Este entorno es el marco dentro del cual el agente de aprendizaje reforzado interactúa, aprende y perfecciona su política de inversión, mediante un sistema de reglas, estados y recompensas claramente definidos. Para el diseño y gestión de este entorno esencial, hemos recurrido a la librería *gym* de OpenAI, que se ha establecido como un estándar de facto en la investigación de aprendizaje reforzado, destacando por su flexibilidad y por ofrecer una extensa biblioteca de entornos predefinidos

gratuitos. A través de la definición de espacios de acción y observación, gym posibilita la simulación de una amplia gama de entornos de decisión. Incluyendo complejos escenarios financieros, facilitando así la adaptación del entorno a las necesidades específicas del proyecto. En nuestro contexto, *gym* juega un papel crucial al simular un mercado financiero en el cual el agente puede tomar decisiones de compra, venta o mantenimiento, basadas en la información de mercado encapsulada en los estados del entorno.

Además, para reforzar la robustez y fiabilidad de la política de inversión estimada por el agente, hemos integrado el uso de *DummyVecEnv* de la librería *stable_baselines3*. Esta herramienta es fundamental para la creación de múltiples entornos de entrenamiento paralelos, actuando como un conjunto de entornos que permiten un entrenamiento simultáneo. La práctica de entrenar al agente en diversas instancias del entorno al mismo tiempo juega un papel significativo en la reducción de la varianza de la estimación del error. Esto resulta en una política de inversión más consistente y resiliente ante datos atípicos, una cualidad de gran valor en el contexto financiero, donde los eventos extremos son frecuentes.

Una vez construido y diseñado el entorno se debe de realizar la **partición de datos, para posteriormente entrenar al agente**. Hemos seleccionado un conjunto de datos de entrenamiento que se extiende desde el 1 de enero de 2008 hasta el 1 de julio de 2022. Este intervalo temporal ha sido para exponer al modelo a una diversidad de escenarios de mercado, incluidas etapas de alta volatilidad, periodos de recesión y ciclos de crecimiento económico. La elección de este amplio rango asegura que el agente esté capacitado para operar en un espectro variado de situaciones de mercado, preparándolo para enfrentar con eficacia los desafíos inherentes a la dinámica del mercado financiero. El entorno de entrenamiento, denominado *StockPortfolioEnv*, recrea un mercado de valores y brinda al agente la posibilidad de interactuar con este de manera cíclica diaria. En cada jornada, el agente evalúa la información del mercado para tomar decisiones de compra, venta o mantenimiento de acciones. A través de un proceso de aprendizaje basado en prueba y error, el agente se familiariza con las estrategias que maximizan el valor de la cartera, mientras que las acciones que resultan en disminuciones son desincentivadas, promoviendo así la optimización de los retornos ajustados por riesgo a lo largo del tiempo.

El propósito central de este experimento es investigar cómo distintas tasas de aprendizaje afectan el desempeño de agentes de Deep Reinforcement Learning en un entorno de mercado. Dada la ausencia de una tasa de aprendizaje óptima universal, reconocemos que su efectividad puede variar significativamente dependiendo de las particularidades del problema a resolver. Por ello, nos proponemos evaluar el impacto que diferentes tasas tienen sobre la eficacia y la eficiencia del aprendizaje de los agentes en alcanzar una política óptima de inversión.

El **diseño experimental** comprende la selección de cinco tasas de aprendizaje distintas [0.0001, 0.01, 0.1, 1, 3] para entrenar 25 agentes por cada tasa. Este enfoque se selecciona con el fin de explorar la sensibilidad de la política de inversión del agente y su proceso de aprendizaje ante variaciones en este hiperparámetro clave. Cada grupo de agentes se entrena durante 200,000 timesteps (días) sobre el conjunto de entrenamiento, lo que permite a los agentes experimentar con un espectro amplio de condiciones de mercado y ajustar sus estrategias de inversión de manera acorde. Posteriormente, el rendimiento de los agentes se evalúa en un conjunto de validación, distinto al de entrenamiento, utilizando el ratio de Sharpe como el principal indicador de rendimiento. Este ratio mide la rentabilidad ajustada al riesgo de la política de inversión, proporcionando un criterio cuantitativo para la evaluación del desempeño de los agentes. Siguiendo los principios de la teoría central del límite, el experimento se estructura para incluir 25 ejecuciones por cada tasa de aprendizaje (Garrido, 2024). Esta metodología nos permite realizar una estimación confiable de la media y la varianza de los ratios de Sharpe obtenidos para cada tasa, estableciendo una comparativa fundamentada sobre el impacto de las tasas de aprendizaje en el rendimiento de los agentes.

5.5. Análisis de resultados

Para la visualización y comparación de los resultados, se utilizarán box plots, que proporcionan una representación gráfica efectiva de la distribución de los ratios de Sharpe para cada tasa de aprendizaje. Esta herramienta estadística resalta la mediana como medida de tendencia central y el rango intercuartílico para evaluar la variabilidad, además de identificar posibles valores atípicos.

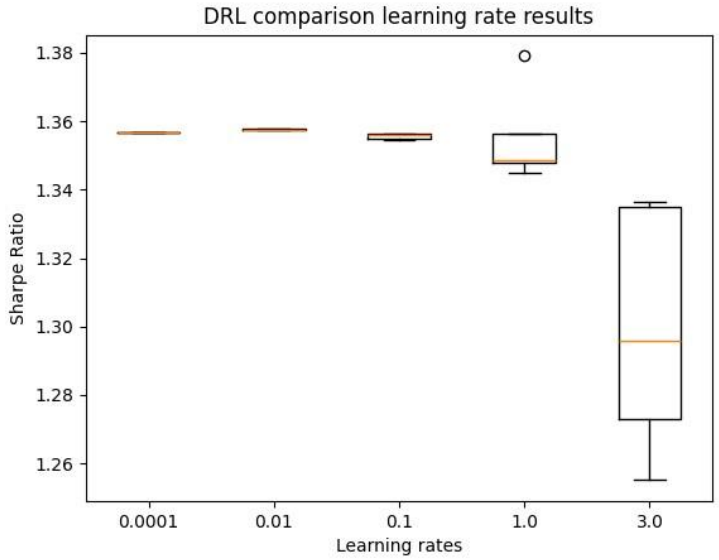


Ilustración 4. Distribución del Ratio de Sharpe para Diferentes Tasas de Aprendizaje

Fuente: Eduardo Garrido Merchán, 2024

Los resultados indican que tasas de aprendizaje más elevadas se asocian con una mayor variabilidad en los ratios de Sharpe, lo que podría interpretarse como una señal de menor estabilidad en las políticas de inversión desarrolladas por los agentes DRL. Es notable una ligera mejora en el rendimiento al incrementar la tasa de aprendizaje de 0.0001 a 0.01. Sin embargo, al superar este umbral, el rendimiento tiende a decrecer con tasas de aprendizaje más altas.

El análisis de la mediana de los ratios de Sharpe revela que no existe una mejora lineal y constante al reducir la tasa de aprendizaje. La tasa de 0.01 se destaca por ofrecer la mediana más alta, lo cual sugiere que este nivel proporciona un equilibrio óptimo, superando tanto la tasa de aprendizaje predeterminada como las tasas más elevadas evaluadas. Este comportamiento sugiere un efecto cuadrático, donde una tasa intermedia es más beneficiosa que una extremadamente alta o baja. Las tasas de aprendizaje elevadas, específicamente 1 y 3, conducen a políticas de inversión más volátiles, como se refleja en la mayor amplitud de los cuartiles y en la frecuencia de valores atípicos. Esto confirma la hipótesis de que altas tasas de aprendizaje pueden provocar reacciones excesivas a las fluctuaciones del mercado, resultando en decisiones de inversión inconsistentes y potencialmente perjudiciales. Por otro lado, tasas más bajas muestran mayor consistencia, aunque no necesariamente se traducen en una mejora sustancial del rendimiento mediano.

Los hallazgos obtenidos permiten descartar la hipótesis nula (H0) que sugiere que no existen diferencias significativas en el rendimiento o que la tasa de aprendizaje predeterminada es igual o más efectiva que una elevada. Además, se confirma parcialmente la hipótesis alternativa

(H1), demostrando que una tasa de aprendizaje alta no mejora necesariamente el rendimiento medido por el ratio de Sharpe y puede degradarlo debido a la inestabilidad que introduce en el proceso de aprendizaje.

El análisis subraya la necesidad de un ajuste meticuloso en la selección de la tasa de aprendizaje para optimizar el desempeño de los algoritmos DRL en el ámbito financiero, que es inherentemente complejo y dinámico. Una tasa intermedia parece ofrecer la mejor sinergia entre estabilidad y rendimiento óptimo, mientras que tasas excesivamente altas comprometen la coherencia de las políticas de inversión y aumentan la incertidumbre en los resultados obtenidos. Investigaciones futuras podrían explorar la identificación de la tasa de aprendizaje óptima y evaluar la robustez de las políticas derivadas frente a variadas condiciones de mercado.

Capítulo 6

Conclusiones y futuras líneas de investigación

Este Trabajo de Fin de Grado ha explorado el impacto de diversas tasas de aprendizaje en la eficiencia de los algoritmos de DRL aplicados a la gestión de carteras. A través de una experimentación rigurosa y un análisis exhaustivo, se ha establecido que la tasa de aprendizaje es un factor crítico en la optimización de políticas de inversión, observándose una mejora en el rendimiento al incrementar la tasa de aprendizaje de 0.0001 a 0.01. No obstante, tasas superiores no garantizan necesariamente mejoras adicionales y pueden, de hecho, incrementar la variabilidad y disminuir la estabilidad de las políticas de inversión. Se ha determinado que una tasa de aprendizaje intermedia de 0.01 ofrece un equilibrio óptimo entre estabilidad y rendimiento eficaz, lo cual es preferible para maximizar el ratio de Sharpe ajustado por riesgo en la gestión de carteras financieras. La evidencia recopilada permite rechazar la hipótesis nula de que no existen diferencias significativas en el rendimiento entre las tasas estudiadas y confirma parcialmente la hipótesis alternativa, sugiriendo que tasas de aprendizaje moderadamente elevadas pueden ser beneficiosas. Estos hallazgos destacan la necesidad de un ajuste meticuloso de la tasa de aprendizaje en la implementación de algoritmos de inversión automatizados y considerando las particularidades del mercado financiero para optimizar la adaptabilidad y la estabilidad del rendimiento.

Una de las vías más prometedoras para extender los hallazgos de este Trabajo de Fin de Grado consiste en explorar la implementación de una estrategia de learning rate dinámico en los algoritmos de DRL. A la luz de los resultados obtenidos, que destacan un rendimiento óptimo con una tasa de aprendizaje intermedia, se propone una investigación futura que evalúe la efectividad de iniciar el proceso de aprendizaje con una tasa relativamente alta. Esto permitiría al algoritmo aproximarse más rápidamente a una política óptima y evitar quedar atrapado en óptimos locales preliminares. Posteriormente, la tasa de aprendizaje se reduciría gradualmente para afinar la política de inversión y alcanzar el óptimo global con mayor precisión. El enfoque propuesto busca ajustar de manera proactiva el learning rate en función de las fases del proceso de aprendizaje: una fase inicial de exploración agresiva seguida por una fase de explotación y optimización detallada. Se hipotetiza que esta metodología dinámica podría superar las limitaciones observadas con tasas fijas, especialmente en entornos de mercado volátiles y complejos donde la adaptabilidad y la capacidad de respuesta rápida son cruciales. El estudio propuesto investigaría de manera exhaustiva cómo diferentes esquemas de ajuste del learning rate afectan la convergencia y la estabilidad del aprendizaje en contextos reales y simulados. Además, se analizaría si un ajuste dinámico del learning rate, de mayor a menor, logra superar estadísticamente el rendimiento de la mejor tasa de aprendizaje fija identificada en este estudio.

Declaración por el uso de la Inteligencia Artificial

Por la presente, yo, Román Martín Gallego, estudiante de E2 Analytics de la Universidad Pontificia Comillas al presentar mi Trabajo Fin de Grado titulado "Influencia del Learning Rate en el Desempeño de Agentes de Deep Reinforcement Learning para Estrategias de Gestión de Carteras", declaro que he utilizado la herramienta de Inteligencia Artificial Generativa ChatGPT u otras similares de IAG de código sólo en el contexto de las actividades descritas a continuación:

1. **Corrector de estilo literario y de lenguaje:** Para mejorar la calidad lingüística y estilística del texto.
2. **Traductor:** Para traducir textos de un lenguaje a otro.

Afirmo que toda la información y contenido presentados en este trabajo son producto de mi investigación y esfuerzo individual, excepto donde se ha indicado lo contrario y se han dado los créditos correspondientes (he incluido las referencias adecuadas en el TFG y he explicitado para que se ha usado ChatGPT u otras herramientas similares). Soy consciente de las implicaciones académicas y éticas de presentar un trabajo no original y acepto las consecuencias de cualquier violación a esta declaración.

Fecha: 22/04/2024

Firma: Román Martín Gallego

Bibliografía

¿Qué es Deep Learning? | IBM. (s. f.).

Ahmed, A., Ghoneim, A., & Saleh, M. (2020). Optimizing stock market execution costs using reinforcement learning. 2020 IEEE Symposium Series on Computational Intelligence (SSCI).

Aitor López Sánchez (2021, julio). Aplicación de Aprendizaje Profundo por Refuerzo a Problemas de Robótica Aérea.

Alonso, M. N., & Srivastava, S. (2020). Deep Reinforcement Learning for Asset Allocation in US Equities. CompSciRN: Other Machine Learning (Topic).

Bajpai, S. (2021). Application of deep reinforcement learning for Indian stock trading automation. ArXiv.

Barreto, S., & Barreto, S. (2024, 18 marzo). Choosing a Learning Rate | Baeldung on Computer Science. Baeldung On Computer Science.

Bartram, S. M., Branke, J., De Rossi, G., & Motahari, M. (2021). Machine Learning for Active Portfolio Management.

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. Journal of Computational Science, 2(1), 1-8.

Brim, A. (2020). Deep Reinforcement Learning Pairs Trading with a Double Deep Q-Network. 2020 10th Annual Computing and Communication Workshop and Conference (CCWC), 0222-0227.

Brunton, S. L., Noack, B. R., & Petros, K. (2019). Machine learning for fluid mechanics. Annual Review of Fluid Mechanics, 52, 477-508.

Buehler, H., et al. (2019).

Buehler, H., Gonon, L., Teichmann, J., & Wood, B. (2019). Deep hedging. Quantitative Finance, 19(8), 1271-1291.

Chen, C.-h., & Zhou, Y. (2021). Application of Deep Reinforcement Learning Algorithm in Smart Finance.

Dean, J. (2014). Big data, data mining, and machine learning: Value creation for business leaders and practitioners. Wiley.

Debie, P., Verhulst, M., Pennings, J., Tekinerdogan, B., Çatal, C., Naumann, A., Demirel, S., Moneta, L., Alskaif, T., Rembser, J., & van Leeuwen, P. (2023). The Analysis of High-Frequency Finance Data using ROOT. Journal of Physics: Conference Series.

Dixon, M., Halperin, I., & Bilokon, P. (2020). Applications of Reinforcement Learning.

Dixon, M., Klabjan, D., & Bang, J. H. (2020). Classification-based financial markets prediction using deep neural networks. Algorithmic Finance, 8(3-4), 147-160.

Dong, Z., Huang, S., Ma, S., & Qian, Y. (2021). Factor Representation and Decision Making in Stock Markets Using Deep Reinforcement Learning. ArXiv.

Dong, Z., Huang, S., Ma, S., & Qian, Y. (2021). Factor Representation and Decision Making in Stock Markets Using Deep Reinforcement Learning. ArXiv,

Erokhin, V., & Lukashenko, I. I. (2022). ECOLOGY AND REGIONAL ENERGY CONSERVATION POLICY. EKONOMIKA I UPRAVLENIE: PROBLEMY, RESHENIYA.

Fazli, M., Lashkari, M., Taherkhani, H., & Habibi, J. (2022). A Novel Experts Advice Aggregation Framework Using Deep Reinforcement Learning for Portfolio Management. ArXiv,

Fiorini, P., & Fiorini, P.-G. (2021). A Simple Reinforcement Learning Algorithm for Stock Trading. 2021 11th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), 2, 824-830.

Floratos, P., Tsantekidis, A., Passalis, N., & Tefas, A. (2022). Online Knowledge Distillation for Financial Timeseries Forecasting. 2022 International Conference on INnovations in Intelligent SysTems and Applications (INISTA), 1-6.

Gao, N., He, Y., Jiao, Y., & Chang, Z. (2021). Online Optimal Investment Portfolio Model Based on Deep Reinforcement Learning. 2021 13th International Conference on Machine Learning and Computing.

Gao, Y., Gao, Z., Hu, Y., Song, S., Jiang, Z., & Su, J. (2021). A Framework of Hierarchical Deep Q-Network for Portfolio Management., 132-140.

Gao, Z., Gao, Y., Hu, Y., Jiang, Z., & Su, J. (2020). Application of Deep Q-Network in Portfolio Management. 2020 5th IEEE International Conference on Big Data Analytics (ICBDA), 268-275.

García Rodriguez, D. (2020). Industrial IoT. Machine Learning en la industria 4.0.

Garrido Merchán, E. C. (2023). RL.

Gašperov, B., & Kostanjčar, Z. (2021). Market Making With Signals Through Deep Reinforcement Learning. IEEE Access, 9, 61611-61622.

Gašperov, B., & Kostanjčar, Z. (2021). Market Making with Signals Through Deep Reinforcement Learning. IEEE Access, 9, 61611-61622.

Gašperov, B., & Kostanjčar, Z. (2022). Deep Reinforcement Learning for Market Making Under a Hawkes Process-Based Limit Order Book Model. IEEE Control Systems Letters, 6, 2485-2490.

Gašperov, B., Begušić, S., Šimović, P. P., & Kostanjčar, Z. (2021). Reinforcement Learning Approaches to Optimal Market Making. Mathematics.

Ghadekar, P., Akolkar, P., Anand, D., Oswal, P., Dixit, S., & Chandak, N. (2022). Mergers and Acquisitions Prediction using Hybrid-Machine Learning and Deep Learning Approach. 2022 IEEE 7th International Conference on Recent Advances and Innovations in Engineering (ICRAIE).

González Oviedo, R. J. (2023). Análisis de dos algoritmos de Reinforcement Learning aplicados a OpenAi Gym Retro: DQN y PPO aplicado entrenamiento de game agents en Ice Climbers. Universidad de Los Andes.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.

Guerra, S. (2023, 14 agosto). La IA en el «trading»: cómo la IA y el aprendizaje automático están cambiando el panorama. lavozdelsur.es.

Guha, S. (2021, 14 diciembre). Deep Deterministic Policy Gradient (DDPG): Theory and implementation. Medium.

Han, Z., Ma, Y., Wang, Y., Günnemann, S., & Tresp, V. (2020). Graph Hawkes Neural Network for Forecasting on Temporal Knowledge Graphs. arXiv: Learning.

Hao, L., Wang, B., Lu, Z., & Hu, K. (2022). Application of Deep Reinforcement Learning in Financial Quantitative Trading. 2022 4th International Conference on Communications, Information System and Computer Engineering (CISCE), 466-471.

Hirsa, A., Hadji Misheva, B., Osterrieder, J., & Posth, J. (2021). Deep Reinforcement Learning on a Multi-Asset Environment for Trading. International Political Economy: Investment & Finance eJournal.

Huang, J., Chang, Q., & Chakraborty, N. (2019). Machine preventive replacement policy for serial production lines based on reinforcement learning. In 15th International Conference on Automation Science and Engineering (CASE) (pp. 1-6). Vancouver.

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. Science, 349(6245), 255-260.

Jose, L. A. J. (2017, 9 junio). CONSTRUCCIÓN DE UN PORTFOLIO DE INVERSIÓN COMPUESTO POR EMISORAS DEL RAMO AUTOMOTRIZ QUE COTIZAN EN LA BMV.

Joshi, D. (2022). Portfolio Optimization using Reinforcement Learning. INTERANTIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT.

Joshi, D. (2022). Portfolio Optimization using Reinforcement Learning. INTERANTIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT.

Joshi, D. (2022). Portfolio Optimization using Reinforcement Learning. INTERANTIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT.

Lee, J., Kim, R., Yi, S., & Kang, J. (2020). MAPS: Multi-Agent reinforcement learning-based Portfolio management System., 4520-4526.

Lee, J., Koh, H., & Choe, H. (2021). Learning to trade in financial time series using high-frequency through wavelet transformation and deep reinforcement learning. Applied Intelligence, 51, 6202 - 6223.

Li, X. (2018). Machine learning in financial markets: A guide to contemporary practice. Quantitative Finance, 18(8), 1315-1329.

Liu, X. Y., Yang, H., Gao, J., & Wang, C. (2021). FinRL: Deep reinforcement learning framework to automate trading in quantitative finance. Proceedings of the Second ACM International Conference on AI in Finance.

Liu, X.-Y., et al. (2021). FinRL: Deep Reinforcement Learning Framework to Automate Trading in Quantitative Finance.

Lommers, K., El Harzli, O., & Kim, J. (2021). Confronting Machine Learning with Financial Data. PSN: Technology (Topic).

Lucarelli, G., & Borrotti, M. (2020). A deep Q-learning portfolio management framework for the cryptocurrency market. Neural Computing and Applications, 32, 17229 - 17244.

Ma, Y., Wang, Z., & Fleiss, A. (2021). Deep Q-Learning for Trading Cryptocurrency.

Markowitz, H. (1952). Portfolio Selection. The Journal of Finance, 7(1), 77-91.

Merchán, E. C. G. (2023, 18 julio). La evolución de la gestión de carteras: Del modelo de Markowitz al Deep Reinforcement Learning y la Optimización Bayesiana.

Meyer, G., Robinson, M., Rasheed, H., & San, O. (2019). Application of Deep Reinforcement Learning using Proximal Policy Optimization on Autonomous Vehicles. Vehicle System Dynamics.

Pawaskar, S. (2022). Stock Price Prediction using Machine Learning Algorithms. International Journal for Research in Applied Science and Engineering Technology.

Pröllochs, N., & Feuerriegel, S. (2018, September 29). Reinforcement Learning in R. ArXiv.

Qi, J., & Ventre, C. (2022). Incentivising Market Making in Financial Markets. Proceedings of the Third ACM International Conference on AI in Finance.

Ruiz Rueda, D. (2021). Aplicación de técnicas de Deep Learning para la gestión de carteras

Sadighian, J. (2019). Deep Reinforcement Learning in Cryptocurrency Market Making. arXiv: Trading and Market Microstructure.

Sánchez, J. (2022). Aprendizaje automático, una revolución para las inversions.

Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). Proximal Policy Optimization Algorithms. ArXiv.

Selser, M., Kreiner, J., & Maurette, M. (2021). Optimal Market Making by Reinforcement Learning. ArXiv.

Selser, M., Kreiner, J., & Maurette, M. (2021). Optimal Market Making by Reinforcement Learning.

Such, I. (2024, 10 de febrero). Agente inversor para acciones de small cap mediante el uso de Reinforcement Learning.

Sun, T., Huang, D., & Yu, J. (2022). Market Making Strategy Optimization via Deep Reinforcement Learning. IEEE Access, 10, 9085-9093.

Ti, Z., Wei Deng, X., & Yang, H. (2020). Wake modeling of wind turbines using machine learning. Applied Energy, 257.

Tripathi, V. (2019). On Present Use of Machine Learning based Automation in Finance.

Trna, M., & Giménez-Martínez, V. (2012). An interactive tool for the stock market research using recursive neural networks. Int. J. Adv. Intell. Paradigms, 4, 103-119.

Tsantekidis, A., Passalis, N., & Tefas, A. (2020). Improving Deep Reinforcement Learning for Financial Trading Using Neural Network Distillation. 2020 IEEE 30th International Workshop on Machine Learning for Signal Processing (MLSP), 1-6.

Tsantekidis, A., Passalis, N., Toufa, A.-S., Saitas-Zarkias, K., Chairistanidis, S., & Tefas, A. (2020). Price Trailing for Financial Trading Using Deep Reinforcement Learning. IEEE Transactions on Neural Networks and Learning Systems, 32, 2837-2846.

Xiao, Q., Li, C., Tang, Y., & Chen, X. (2020). Energy efficiency modeling for configuration-dependent machining via machine learning: A comparative study. IEEE Transactions on Automation Science and Engineering, 1-14.

Zhu, H. (2022). Researches advanced in financial trading systems based on reinforcement learning.

Zhu, Y., Wang, L., & Zhang, Y. (2020). Proximal Policy Optimization for Workflow Management in Manufacturing Systems. IEEE Transactions on Industrial Informatics. https://doi.org/10.1109/TII.2020.2976706

Zvornicanin, E., & Zvornicanin, E. (2024b, marzo 18). Relation Between Learning Rate and Batch Size | Baeldung on Computer Science. Baeldung On Computer Science.