Gestión activa de portafolios de cripto activos utilizando técnicas de aprendizaje por refuerzo

Presentado para obtener el título de:

MAGÍSTER EN MATEMÁTICAS APLICADAS Y CIENCIAS DE LA COMPUTACIÓN

Emilio Muñoz Pérez

Dirección:

Edgar J. Andrade-Lotero

Universidad del Rosario

Escuela de Ingeniería, Ciencia y Tecnología

Maestría en matemáticas aplicadas y ciencias de la computación

AGRADECIMIENTOS

Quiero expresar mi más sincero agradecimiento a Edgar Andrade por su inmenso apoyo y dedicación durante todo el proceso de creación de mi tesis. Su disposición para orientarme, su paciencia y su pasión por el Reinforcement Learning han sido fundamentales para mi desarrollo académico y personal. Gracias a su inspiración y orientación, he descubierto una verdadera pasión por este campo, y eso es lo más valioso que un estudiante puede recibir de sus docentes.

ABSTRACT

Español

En un entorno financiero marcado por la volatilidad y la falta de transparencia que caracteriza al mercado de criptoactivos, la gestión de portafolios se enfrenta a desafíos significativos. Tradicionalmente, las estrategias de gestión de activos se ven limitadas por la impredecibilidad de este sector en constante evolución. Este estudio se propone abordar este desafío mediante la aplicación del aprendizaje por refuerzo, una técnica de aprendizaje automático que utiliza la retroalimentación de un agente para aprender y adaptarse de manera continua. En este contexto, el "agente" es el portafolio de criptoactivos y las "recompensas" son los retornos financieros que este logra obtener.

El objetivo de este enfoque es permitir que el portafolio aprenda de la retroalimentación en tiempo real que proviene del mercado de criptoactivos y, en consecuencia, ajuste de manera continua la asignación de activos. Esto se realiza con la finalidad de maximizar el rendimiento del portafolio y superar las estrategias de inversión pasiva en activos digitales. A través del aprendizaje por refuerzo, se espera que el portafolio se adapte de manera eficiente a los cambios del mercado y tome decisiones óptimas para mejorar los retornos y minimizar el riesgo.

Para evaluar la efectividad de este enfoque, se utilizarán datos históricos de precios de criptoactivos. El modelo basado en aprendizaje por refuerzo se comparará con otras estrategias de gestión de portafolios, como la asignación pasiva de activos. El resultado principal obtenido es que el modelo por refuerzo tiene un desempeño superior, generando mejores rendimientos y menor volatilidad en comparación con las estrategias tradicionales.

En resumen, este trabajo busca demostrar que el aprendizaje por refuerzo puede ser una técnica efectiva para mejorar la gestión de portafolios de criptoactivos. Al adaptarse de manera dinámica a las condiciones cambiantes del mercado, este enfoque permite una optimización continua de la

asignación de activos, maximizando el retorno y reduciendo el riesgo. En un escenario de inversiones digitales en constante evolución, esta investigación ofrece una perspectiva prometedora para quienes buscan gestionar activos de manera eficiente en el mundo de los criptoactivos.

English

In a financial environment characterized by volatility and the lack of transparency that defines the crypto-asset market, portfolio management faces significant challenges. Traditionally, asset management strategies are limited by the unpredictability of this constantly evolving sector. This study aims to address this challenge by applying reinforcement learning, a machine learning technique that uses an agent's feedback to learn and adapt continuously. In this context, the "agent" is the crypto-asset portfolio, and the "rewards" are the financial returns it manages to obtain.

The goal of this approach is to allow the portfolio to learn from real-time feedback from the crypto-asset market and, consequently, continuously adjust asset allocation. This is done to maximize portfolio performance and outperform passive investment strategies in digital assets. Through reinforcement learning, the portfolio is expected to efficiently adapt to market changes and make optimal decisions to improve returns and minimize risk.

To assess the effectiveness of this approach, historical crypto-asset price data will be used. The reinforcement learning-based model will be compared with other portfolio management strategies, such as passive asset allocation. The main outcome is that the reinforcement model performs better, generating higher returns and lower volatility compared to traditional strategies.

In summary, this work seeks to demonstrate that reinforcement learning can be an effective technique for improving crypto-asset portfolio management. By dynamically adapting to changing market conditions, this approach allows for continuous optimization of asset allocation, maximizing return and reducing risk. In a constantly evolving digital investment scenario, this

research offers a promising perspective for those looking to manage assets efficiently in the world of crypto-assets.

Tabla de contenido

Capítulo 1
INTRODUCCIÓN

La presente investigación aborda la optimización en la gestión de portafolios de criptoactivos, empleando técnicas avanzadas de aprendizaje por refuerzo. La gestión de portafolios se refiere al proceso de asignación y reajuste de un conjunto de activos financieros, como acciones o criptomonedas, con el objetivo de maximizar los retornos esperados y minimizar los riesgos asociados (Goodfellow, Bengio, & Courville, 2016). Los criptoactivos, en particular, han ganado una notable prominencia en años recientes, emergiendo como una nueva categoría de activos financieros. Su creciente popularidad ha llevado a fondos de cobertura y otras entidades financieras a considerar su inclusión en sus estrategias de inversión (Elendner, Trimborn, Ong, & Lee, 2018; Liu & Tsyvinski, 2018).

No obstante, la gestión de estos activos se ve desafiada por la volatilidad de sus precios y la falta de transparencia en un mercado no regulado, donde la información no está equitativamente distribuida entre todos los participantes. Adicionalmente, la gestión de un portafolio se ha abordado mediante técnicas de optimización convencionales y análisis estadístico. Sin embargo, la eficacia de estos métodos es limitada debido a la naturaleza dinámica y compleja de los mercados de criptoactivos.

En este contexto, se propone superar la gestión pasiva de portafolios de criptoactivos mediante la aplicación de técnicas avanzadas de aprendizaje por refuerzo. Este enfoque se espera que logre una optimización más efectiva, maximizando rendimientos y siendo menos susceptible a las tendencias alcistas o bajistas del mercado. Basándonos en la metodología de aprendizaje automático, en la que un agente interactúa con un entorno para aprender a tomar decisiones óptimas (Sutton & Barto, 2018), se entrenará un agente específicamente para la gestión de portafolios de criptoactivos. Este agente se capacitará

para tomar decisiones de inversión óptimas, apoyándose en un análisis detallado del rendimiento histórico del mercado y las fluctuaciones de precios.

Esto se logra mediante la incorporación de indicadores financieros, que definen el "espacio de estados" en el que operará el agente. En el aprendizaje por refuerzo, el "espacio de estados" comprende todas las situaciones posibles que el agente puede observar o experimentar. En el contexto de la gestión de portafolios de criptoactivos, este espacio incluiría variables como los precios de los criptoactivos, volúmenes de mercado, indicadores técnicos, y posiblemente señales empleadas por los profesionales del sector. Estos elementos proporcionarán al agente un marco integral para la toma de decisiones informadas y efectivas.

El modelo propuesto será entrenado y evaluado utilizando datos históricos del mercado, mediante técnicas de backtesting, y se comparará su rendimiento con una gestión pasiva. Los resultados obtenidos no solo soportan la tesis de que un agente basado en aprendizaje por refuerzo gestiona mejor un portafolio de criptoactivos que las estrategias pasivas tradicionales, sino que también permiten dar luces sobre mejores estrategias de inversión basadas en la política seguida por el agente.

Capítulo 2
OBJETIVOS

Para responder la hipótesis se planearon los siguientes objetivos:

Objetivo general

- Explorar el uso de aprendizaje por refuerzo en la administración de portafolio.

Objetivos específicos

- Estudio general del mercado de cripto activos y selección de monedas relevantes para una estrategia de gestión activa de portafolio.
- Definición componentes del modelo; estados, acciones y ambiente.
- Prueba de aprendizaje por refuerzo en un entorno controlado.
- Evaluación y comparación de los resultados del análisis.

Capítulo 3
PROBLEMA Y JUSTIFICACIÓN

La problemática radica en que el mercado de criptoactivos es intrínsecamente volátil y opaco, con precios que pueden fluctuar drásticamente en cuestión de minutos debido a factores que abarcan desde la especulación hasta intervenciones regulatorias (Jing & Rocha, 2023). La gestión de portafolios en este contexto se ve comprometida por la imprevisibilidad y la falta de transparencia, lo que dificulta la aplicación de estrategias de inversión tradicionales. Además, la inmadurez del mercado de criptoactivos, en comparación con los mercados financieros tradicionales, presenta un menor número de herramientas y referentes confiables para que los inversores puedan medir el desempeño y el riesgo de sus inversiones. Estas características suponen un reto significativo para inversores y gestores de fondos que buscan optimizar sus retornos mientras minimizan los riesgos asociados.

En el contexto descrito, este trabajo adquiere una importancia crítica por varias razones. Hay una relevancia técnica en probar cómo la aplicación del aprendizaje por refuerzo en la gestión de portafolios de criptoactivos representa un avance significativo. Esto permite superar las limitaciones de las estrategias pasivas y adaptativas convencionales mediante la utilización de algoritmos que aprenden y se ajustan en tiempo real a la dinámica del mercado.

Por otro lado, puede generar un impacto económico, ya que la eficiencia en la gestión de portafolios de criptoactivos tiene el potencial de mejorar los rendimientos de las inversiones. Al maximizar los retornos y reducir la volatilidad, los inversores pueden obtener beneficios más consistentes, lo que es crucial en un entorno de inversión que está creciendo en tamaño y complejidad. Esto también denota una aplicabilidad práctica, ya que esta investigación puede proporcionar un marco replicable o adaptable en diferentes

contextos de inversión, ampliando su aplicabilidad más allá del mercado de criptoactivos e influyendo en la gestión de portafolios tradicionales.

Finalmente, a medida que el mercado de criptoactivos se expande, se espera que las políticas y regulaciones financieras evolucionen para adaptarse a esta nueva clase de activos. Un modelo de gestión de portafolios robusto y basado en aprendizaje automático puede proporcionar una base sólida para el desarrollo de políticas informadas y efectivas.

Por lo tanto, este trabajo se posiciona en la intersección de la innovación tecnológica y la necesidad práctica de soluciones de inversión para un abanico de nuevos activos digitales. La solución propuesta no solo busca abordar un problema económico inmediato, sino también contribuir a la construcción de un marco más resiliente y adaptativo para la gestión financiera en el ámbito digital.

Capítulo 4
MARCO TEÓRICO Y ESTADO DEL ARTE

I. Marco teórico

Hay tres elementos claves que deben ser descritos para entender el contexto en el que se desarrolla el presente trabajo de grado y serán descritos a continuación:

Criptomonedas

Las criptomonedas son activos digitales diseñados para funcionar como un medio de intercambio que utiliza la criptografía para asegurar las transacciones, controlar la creación de unidades adicionales y verificar la transferencia de activos. Bitcoin, creado bajo el seudónimo de Satoshi Nakamoto, es la primera y más conocida criptomoneda, lanzada en 2009. Desde entonces, miles de variantes han sido creadas, conocidas como altcoins.

Las criptomonedas se caracterizan por ser descentralizadas; no están respaldadas ni controladas por ningún gobierno o entidad central. Esto las hace inherentemente resistentes a la manipulación o interferencia gubernamental. Operan en una tecnología de libro mayor distribuido llamada blockchain, que es un registro público y compartido de todas las transacciones que se actualiza y mantiene por una red de computadoras.

Gestión de portafolios activa

La gestión de portafolio activa es un enfoque dinámico en el ámbito financiero que busca superar los rendimientos del mercado mediante la selección activa de activos. A diferencia de la optimización de portafolios tradicional, que se centra en la asignación estática de recursos, la gestión de portafolio activa implica tomar decisiones continuas y

basadas en el análisis constante de datos y condiciones del mercado, lo que es un escenario ideal para algoritmos de aprendizaje por refuerzo.

Este enfoque se vuelve especialmente relevante en el contexto de las criptomonedas, donde la volatilidad es elevada y la correlación con otros activos tradicionales es limitada. La gestión de portafolio activa permite a los inversores adaptarse ágilmente a las fluctuaciones del mercado y aprovechar oportunidades emergentes. Se basa en estrategias informadas por el análisis de datos en tiempo real y la aplicación de algoritmos de aprendizaje automático, como se ha explorado en investigaciones recientes (Jiang et al., 2016; Huang et al., 2020; Li et al., 2019; Wang et al., 2020).

Aprendizaje por Refuerzo

El aprendizaje por refuerzo (aprendizaje por refuerzo) es un tipo de aprendizaje automático en el que un agente aprende a tomar decisiones mediante la experimentación en un entorno interactivo. Este agente realiza acciones y recibe recompensas o penalizaciones en forma de señales de retorno, buscando, en última instancia, aprender una estrategia (o "política") que maximice las recompensas futuras (Sutton & Barto, 2018). Las acciones son las distintas opciones que tiene un agente que lo llevaran de un estado inicial a un estado final. Por ejemplo, en un juego de ajedrez una acción es el movimiento de una pieza que hace que el tablero obtenga un estado nuevo (una configuración distinta de fichas). Por último, las recompensas son el elemento que permite determinar si un agente está tomando las acciones adecuadas. Volviendo al caso del ajedrez, serían cruciales para evaluar la efectividad de un movimiento; por ejemplo, si ese movimiento resulta en la pérdida de fichas propias o en la captura de fichas del jugador contrario. Lo que el agente quiere mejorar es la política, que hace referencia a un manual de instrucciones que le indican que acción tomar en cada estado.

Es importante contemplar, que, en aprendizaje por refuerzo, los problemas se modelan como procesos de decisión de Markov, donde las decisiones se toman en una secuencia de estados que dependen solo del estado actual y no de la secuencia de eventos que precedieron. Esto es particularmente útil en la gestión de portafolios de criptomonedas, donde el estado del mercado puede cambiar rápidamente y de maneras impredecibles.

Para aprender la política óptima existe SARSA y Q-learning, que son métodos de control en aprendizaje por refuerzo. SARSA es un algoritmo "on-policy" que actualiza su valor Q basándose en la acción realizada, mientras que Q-learning es un algoritmo "off-policy" que actualiza sus valores Q basándose en la acción óptima estimada (Watkins & Dayan, 1992; Sutton & Barto, 2018). Estos algoritmos resultan particularmente adecuados para la gestión de carteras de criptoactivos debido a su capacidad para manejar la naturaleza estocástica y la incertidumbre del mercado, adaptándose a cambios en tiempo real y aprendiendo de las consecuencias de las acciones tomadas (Jiang, Xu, & Liang, 2017; Li & Hoi, 2014).

En cada paso t de un episodio, el entorno se encuentra en un estado s_t en el caso de gestión de carteras de criptoactivos, el estado puede concebirse como una combinación de la distribución del portafolio y señales del mercado. (ver sección II. Definición del modelo En página 19 Para el modelo concreto del caso). En este estado, el agente ejecuta una acción a_t . que consiste en la asignación porcentual de capital invertido para cada uno de los activos del portafolio. El entorno recibe la acción y devuelve un nuevo estado s_{t+1} y una recompensa r_{t+1} . De acuerdo a la regla de aprendizaje de SARSA, el agente debe actualizar la estimación del valor q(s_t , a_t) así:

$$Q(s,a) \leftarrow Q(s,a) + \alpha(t + \gamma(Qs', a') - Q(s,a))$$

Mientras, que la regla Q-learning tiene en cuenta el propósito de buscar el valor óptimo dentro de la misma actualización, y se define por:

$$Q(s,a) \leftarrow Q(s,a) + \alpha(t + \gamma \underset{a'}{\overset{máx}{}}(Qs', a') - Q(s,a))$$

Integración de los Conceptos

La integración de estos tres conceptos se centra en cómo el aprendizaje por refuerzo puede ser aplicado para la gestión de portafolios de criptomonedas. Dado que las criptomonedas son activos altamente especulativos y volátiles, los métodos tradicionales de optimización de portafolios pueden no ser suficientes. Aquí es donde el aprendizaje por refuerzo puede ofrecer una ventaja significativa, ya que puede adaptarse a entornos inciertos y aprender de la experiencia para mejorar las decisiones de inversión en tiempo real.

El aprendizaje por refuerzo permite que los modelos de gestión de portafolios consideren la maximización de retornos ajustados al riesgo a través de la exploración y explotación de estrategias de inversión en el mercado de criptomonedas. Con el tiempo, el agente de aprendizaje por refuerzo puede identificar patrones en los datos de precios y volumen de las criptomonedas y ajustar las asignaciones de portafolio para navegar en un mercado en constante cambio, buscando maximizar los retornos a largo plazo mientras se gestiona el riesgo asociado (Jiang et al., 2016; Huang et al., 2020; Li et al., 2019; Wang et al., 2020).

II. Estado del arte

La aplicación de técnicas de aprendizaje automático, en particular el aprendizaje por refuerzo (aprendizaje por refuerzo, ha emergido como una metodología prometedora para enfrentar desafíos específicos. El aprendizaje por refuerzo se basa en la premisa de que un agente toma decisiones en un entorno para maximizar una recompensa acumulativa,

según Sutton y Barto (2018). En la gestión de portafolios de criptoactivos, dicho agente es un algoritmo que busca optimizar el retorno financiero.

Existen contribuciones notables en este campo. Por ejemplo, Jiang et al. (2016) desarrollaron uno de los primeros modelos que emplean redes neuronales convolucionales, entrenadas con datos históricos de precios de criptomonedas, para determinar la asignación de un portafolio. Este modelo, que busca maximizar el retorno acumulativo, mostró resultados alentadores en pruebas de backtesting. Posteriormente, Jiang, Xu y Liang (2017) extendieron este enfoque mediante un marco de aprendizaje por refuerzo que utiliza una topología de Ensemble of Identical Independent Evaluators (EIIE), superando otras estrategias en backtesting en el mercado de criptomonedas.

Huang, Zhou y Song (2020) innovaron al incorporar estrategias de venta en corto y arbitraje en su modelo de aprendizaje por refuerzo para la gestión de portafolios, optimizando decisiones de inversión y generando retornos superiores en el mercado de valores. Asimismo, Li, Wang y Cao (2019) abordaron limitaciones prácticas de investigaciones anteriores, como los costos de transacción y las restricciones de venta en corto, proponiendo un marco de aprendizaje por refuerzo para la gestión de activos con pesos de activos continuos y toma de decisiones basadas en características relevantes.

En otro desarrollo significativo, Wang et al. (2020) propusieron un sistema de trading de acciones reforzado jerárquico para la gestión de portafolios, que descompone el proceso de trading en una jerarquía de tareas, logrando mejoras sustanciales en comparación con métodos anteriores.

Además de los progresos mencionados, persisten brechas significativas en la literatura. Muchos modelos existentes no consideran adecuadamente la dinámica en tiempo real del mercado de criptoactivos, ni la capacidad de ajustar la asignación de activos de forma continua ante cambios del mercado en un contexto de gestión de portafolio. Esto es

particularmente crítico en escenarios de alta volatilidad y en períodos de crisis, los cuales son cada vez más habituales en el ecosistema de criptomonedas, como ilustran casos como Luna-Terra, Three Arrows Capital y FTX. Además, otro aspecto poco explorado es la integración efectiva de señales de mercado externas, como indicadores de profesionales del sector, indicadores económicos o eventos geopolíticos, que pueden tener un impacto significativo en los criptoactivos. Este trabajo tendría la versatilidad de poder incluir ese tipo de variables fácilmente.

Consciente de su alcance más limitado comparado con investigaciones anteriores, este documento busca aportar a estas áreas aún inexploradas. La intención es desarrollar un modelo de aprendizaje por refuerzo simplificado, pero eficiente, que pueda adaptarse en tiempo real a las condiciones del mercado. Aunque no aspire a la complejidad de los modelos previos, se enfoca en la maximización de retornos y la minimización de riesgos en el portafolio de criptoactivos, considerando factores clave como la volatilidad (precio máximo y mínimo durante la unidad de tiempo), el volumen tranzado y otro tipo de señales relevantes como lo son los promedios móviles simples. Este enfoque pragmático y más accesible proporcionará una valiosa base para futuras investigaciones y aplicaciones prácticas en el campo de la gestión de portafolios de criptoactivos. Además, una ventaja es que el método resulta más transparente, ya que facilita la exploración de la política del agente, permitiendo así comprender mejor su comportamiento y ofrecer explicaciones sobre sus decisiones a los inversores.

Capítulo 5
METODOLOGÍA

La metodología de esta investigación se centra en el desarrollo y la implementación de un modelo de aprendizaje por refuerzo para la gestión de portafolios de criptoactivos. La metodología se divide en varias fases clave que se describen a continuación:

Recopilación de Datos y Selección de Criptoactivos

1. Recopilación de Datos Históricos:
    - Se obtendrán datos históricos de precios, volumen y otras métricas relevantes de fuentes confiables y públicamente accesibles. Se utilizarán datos de cierre por hora en todos los escenarios, es decir, cada unidad de tiempo será el cierre final del indicador en una hora específica. Esta información se analizará de manera secuencial, ya que se trata de una serie temporal. Los datos se normalizarán y se transformarán para garantizar la consistencia y la calidad necesarias para el análisis.
2. Identificación de Criptoactivos Relevantes:
    - Se realizará un análisis del mercado para identificar las criptomonedas con mayor capitalización de mercado y liquidez.
    - Se seleccionarán cinco criptomonedas basándose en indicadores que las hagan relevantes para una gestión activa de portafolio.

Diseño del modelo

1. Diseño del Modelo y Arquitectura:
    - Se diseñará un modelo de aprendizaje por refuerzo que pueda adaptarse a las condiciones cambiantes del mercado de criptoactivos.

- La arquitectura del modelo incluirá definiciones claras del espacio de estados, las acciones posibles y la política de aprendizaje.

2. Métodos de prueba del Modelos:
    - Primero se ejecutará el modelo en un ambiente con una señal fabricada que permita identificar si el agente efectivamente está aprendiendo. También, solo se usarán dos monedas para Facilitar el entrenamiento
    - Se evaluará si el modelo logró obtener la política perfecta que se identifica para esta prueba reducida y basado en el comportamiento de la recompensa se seleccionará el tipo de agente a utilizar.

Implementación y Entrenamiento del Modelo

Una vez identificado el tipo de modelo a utilizar se entrenará en un entorno donde cuente con más estados y con los cinco activos seleccionados. Para esto es necesario:

1. Preparación de Datos:
    - Los datos históricos se procesarán y estructurarán para su uso en el entrenamiento del modelo. Para esto es necesario determinar las variables que serán las señales que utilice el modelo para aprender, conocidas como características.
    - Se realizará una división de los datos en conjuntos de entrenamiento, prueba.
2. Definición del Espacio de Estados:
    - Se especificará el espacio de estados basado en las características financieras de las criptomonedas seleccionadas.
3. Entrenamiento del Modelo:
    - Se implementará el modelo y se ajustará mediante la iteración sobre distintos hiper parámetros y políticas de enfriamiento.

- Se correrán episodios hasta que el agente no pueda mejorar significativamente el crecimiento de su función de recompensa.

Evaluación y Comparación de Resultados

1. Pruebas y Backtesting:
    - Primero, se llevará a cabo un backtesting, una prueba con datos de otro período de tiempo no utilizado en el entrenamiento. En un escenario, se simulará un mercado con una tendencia alcista, caracterizado por un aumento sostenido de los precios entre la fecha inicial y final del periodo evaluado. En un segundo escenario, se simulará un mercado con una tendencia bajista, marcado por una disminución prolongada de los precios entre dos fechas seleccionadas. Esto es relevante para determinar si el modelo tiene un comportamiento favorable independientemente de la tendencia del mercado.
    - Se comparará el rendimiento del modelo con una estrategia pasiva de gestión de portafolios. En una estrategia pasiva, los inversionistas buscan replicar el desempeño de un índice de mercado en lugar de realizar selecciones activas de inversiones. Esto implica mantener un portafolio diversificado que refleje la composición del índice subyacente, sin realizar cambios en respuesta a movimientos del mercado.

2. Análisis de la política:
    - Se revisará la política del agente para entender que señales está utilizando para tomar decisiones y, en lo posible, entender cuál es el motivo de su desempeño.

Capítulo 6
RESULTADOS Y DISCUSIÓN

I. Estudio general del mercado de cripto activos y selección de monedas relevantes

Se evaluaron 15 activos del mercado de criptoactivos, que representan los de mayor capitalización y adopción de entre más de 350 activos disponibles en la plataforma de intercambio Binance. Además, se excluyeron las StableCoins y los distintos derivados de otras monedas ya existentes en la lista, como el Wrapped BTC, con el propósito de identificar aquellos que resulten más relevantes basados en criterios financieros establecidos. La información descrita a continuación se extrajo del API de Binance el 07 de noviembre de 2023.

Tabla 1 – Criptomonedas y capitalización del mercado (Market Cap)

Precios en (USD $)

Nombre	Ticker	Precio	Market Cap
Bitcoin	BTC	$34,663.58	$677,035,240,782
Ethereum	ETH	$1,876.17	$225,630,254,212
BNB	BNB	$247.50	$37,550,050,778
Ripple	XRP	$0.6838	$36,663,117,617
Solana	SOL	$41.14	$17,308,092,171
Cardano	ADA	$0.3462	$12,200,809,716
Dogecoin	DOGE	$0.07365	$10,438,301,127
TRON	TRX	$0.09654	$8,566,941,745
Toncoin	TON	$2.41	$8,259,797,897
Chainlink	LINK	$12.82	$7,136,655,367
Polygon	MATIC	$0.7118	$6,580,288,392
Polkadot	DOT	$4.88	$6,135,653,691

Litecoin	LTC	$72.85	$5,379,633,453
Dai	DAI	$0.9996	$5,345,565,483
Shiba Inu	SHIB	$0.000008224	$4,846,758,460

Adicional se extrajo y se procesó un conjunto de datos de cada uno de los activos del API de Binance que posee la siguiente estructura, esta información se extrajo por hora:

Tabla 2 – Ejemplo del conjunto de datos procesado de una criptomoneda
Precios en USD$ [open, high, low, close]

date_open	open	high	low	close	volume	date_close
2020-09-01 19:00:00	11983.56	11989.82	11953.05	11957.41	1991.207868	2020-09-01 20:00:00
2020-09-01 20:00:00	11957.41	12015.11	11938.86	12003.60	2365.851556	2020-09-01 21:00:00
2020-09-01 21:00:00	12003.61	12041.86	11988.01	12016.05	2258.630437	2020-09-01 22:00:00
2020-09-01 22:00:00	12016.14	12050.00	11941.80	11973.04	3017.157538	2020-09-01 23:00:00
2020-09-01 23:00:00	11973.04	12011.60	11851.00	11921.97	3388.050415	2020-09-02 00:00:00

Finalmente, una vez identificados estos activos, se realiza un análisis basado en indicadores financieros relevantes para la gestión activa de portafolios, con el fin de identificar cinco activos para que el algoritmo opere. Los criterios utilizados se describen a continuación.

Volatilidad:

La volatilidad representa una medida estadística de la dispersión de rendimientos para un activo, es comúnmente calculada a través de la desviación estándar de los retornos históricos. Esta métrica es de vital importancia, ya que permite a los inversores estimar el riesgo inherente y las expectativas de retorno. En un mercado caracterizado por sus fluctuaciones pronunciadas, como el de las criptomonedas, una estrategia de selección

que privilegie activos con menor volatilidad podría ser una táctica prudente para mitigar riesgos, sin obviar la posibilidad de rendimientos competitivos (Liu & Tsyvinski, 2018).

Volumen:

El volumen promedio operado es un indicador que cuantifica el número total de unidades de criptoactivos intercambiadas durante un periodo establecido. Un volumen elevado suele ser sinónimo de alta liquidez, lo que facilita la ejecución de órdenes de compra o venta sin provocar deslizamientos significativos en el precio. La liquidez es un componente esencial para inversores y traders que buscan adaptabilidad y eficiencia en la gestión de sus inversiones dentro de los mercados volátiles de criptomonedas (Steven Gordon, Zhi Li, John Marthinsen, 2023).

Rango promedio diario:

El rango promedio diario se define como la diferencia entre el precio alto y bajo de un criptoactivo dentro de una jornada de negociación. Este parámetro ofrece una perspectiva más detallada de la volatilidad intradía y puede ser indicativo de oportunidades de arbitraje o especulación basadas en movimientos de precios intradía (Van Heerden et al., 2021).

Precio Promedio Ponderado por Volumen (VWAP):

"Volume Weighted Average Price" (Precio Promedio Ponderado por Volumen), es una métrica financiera que calcula el promedio del precio de un activo, ponderado por su volumen de transacciones. Esta métrica es ampliamente utilizada en varias clases de activos para determinar el valor justo de mercado, proporcionando una referencia para los traders y los inversores sobre el precio promedio al que un activo se ha negociado a lo largo de un período específico. En el contexto de los mercados de criptomonedas, el VWAP juega un papel crucial en la comprensión de la formación de precios y la

ejecución óptima de órdenes. Un estudio reciente de (Li et al. 2022) aborda la optimización de estrategias de VWAP utilizando aprendizaje profundo y aprendizaje por refuerzo jerárquico.

Estis indicadores, aplicados de manera conjunta, posibilita la identificación de criptoactivos que se caracterizan por su estabilidad, liquidez y valoración ajustada al mercado, atributos generalmente buscados por inversores para la conformación de un portafolio de inversión. Finalmente se seleccionan los activos basados en los de menor volatilidad, mayor volumen y rango promedio diario, obteniendo los resultados a continuación:

Tabla 3 – Indicadores financieros calculados por criptomonedas
Precios en USD$ [Rango promedio, VWAP]

Nombre	**Volatilidad**	**Volumen promedio**	**Rango promedio**	**VWAP promedio**
BTC	0.0046	4,647	154.8934	24273.9
ETH	0.0049	15,456	10.7991	1723.79
BNB	0.0049	17,840	1.7173	267.89
TRX	0.005	14,034,343	0.0004	0.07
DOT	0.0065	158,367	0.0475	5.55

* VWAP: Volume Weighted Average Price

Una vez seleccionados los activos, procedimos a la creación de un modelo de aprendizaje por refuerzo.

II. Definición del modelo

A continuación, se presenta una descripción del modelo, su desarrollo y sus componentes.

En el ámbito de la gestión de portafolios de criptoactivos, la aplicación del aprendizaje por refuerzo (aprendizaje por refuerzo) permite el desarrollo de estrategias de inversión automatizadas. El objetivo del agente de aprendizaje por refuerzo es maximizar el retorno de la inversión, tomando decisiones basadas en el análisis de cuándo comprar, vender o mantener posiciones en diferentes criptoactivos. Este enfoque implica la creación de un modelo y sus componentes para evaluar la efectividad de esta tecnología en dicho entorno.

Dada la importancia de precisar los elementos fundamentales del aprendizaje por refuerzo al definir un modelo, realizaremos un breve resumen de estos. En primer lugar, las acciones disponibles para este modelo consistían en la asignación de un porcentaje a cada uno de los activos disponibles, lo que refleja la distribución de activos que tendrá el portafolio. Por ejemplo, una acción podría ser asignar el 0% a todos los activos, indicando que se vendieron todos los activos y el portafolio permanece completamente en USD. En segundo lugar, el estado hace referencia a diversas señales provenientes de los indicadores de mercado y la posición actual del portafolio. Finalmente, la función de recompensa se define como el cambio en el valor de venta de todo el portafolio con respecto al periodo anterior.

La definición formal del entorno que hemos usado es la siguiente:

- Estados: $\langle allocation_{c_1}, \dots, allocation_{c_k} \rangle_t \cup \langle discP_{c_1}, \dots, disc_{c_k} \rangle_t$ donde $allocation_c$ es el porcentaje del capital total invertido en la criptomoneda c; y $discP_c$ es la discretización del cambio porcentual C de la diferencia de los

promedios de la ventana móvil de 5 horas ($\overline{p_c^5}$), 8 horas ($\overline{p_c^8}$) y 13 horas ($\overline{p_c^{13}}$) finalizando en el tiempo t para la criptomoneda c:

$$C = cambio\_porcentual\left(\left|\overline{p_c^5} - \overline{p_c^8}\right| + \left|\overline{p_c^5} - \overline{p_c^{13}}\right| + \left|\overline{p_c^8} - \overline{p_c^{13}}\right|\right)$$

La discretización se hace encontrando el valor $\frac{n}{20}$ del intervalo $\left[\frac{n}{20}, \frac{n+1}{20}\right)$ en el que cae C.

- Acciones: $\langle allocation_{c_1}, \ldots, allocation_{c_k} \rangle$ donde $allocation_c$ es el porcentaje del capital total invertido en la criptomoneda c.
- Recompensa: $p_t - p_{t-1}$, donde $p_t = \sum_c allocation_c * P_c$, donde P_c es el precio de la criptomoneda c en el tiempo t.

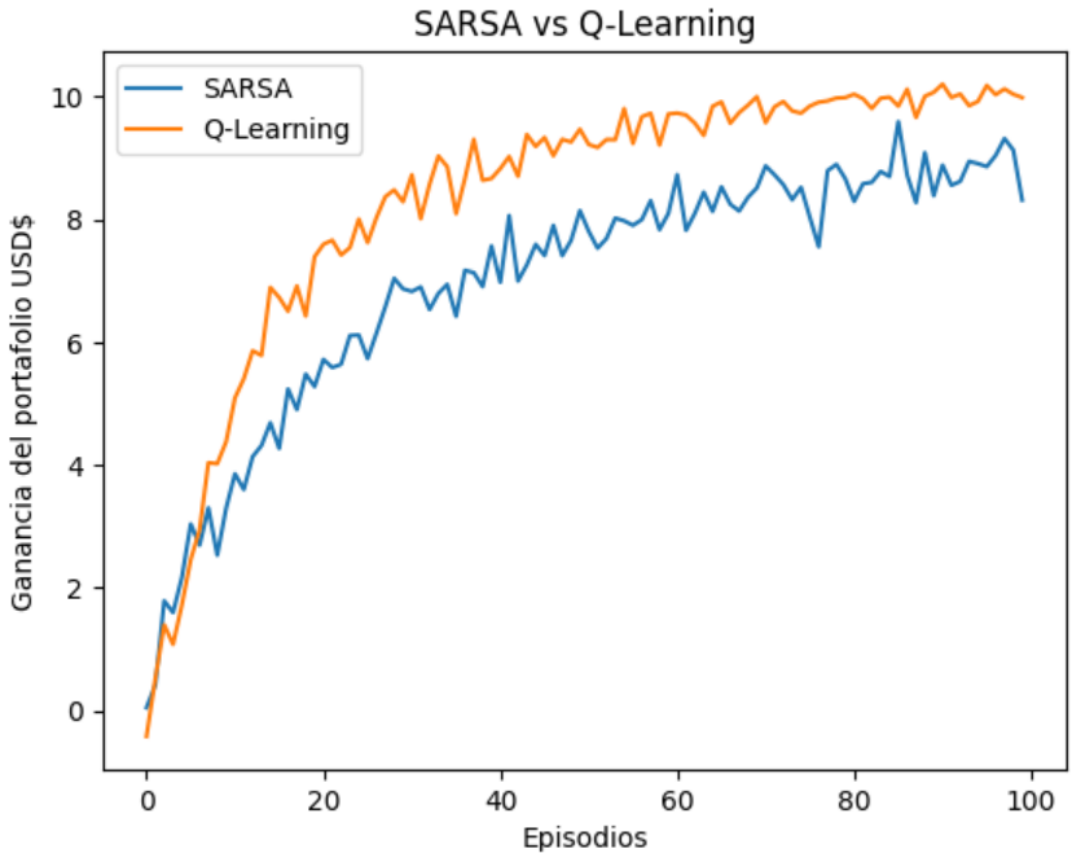
Para la implementación del entorno empleamos la librería Gymnasium de OpenAI, que ya dispone de un entorno predefinido para un solo activo. Creamos una subclase del entorno de Gym para reflejar métricas financieras relevantes, definir acciones de compra, venta o mantenimiento, y establecer una estructura de recompensas basada en el rendimiento del portafolio y otros indicadores financieros. Además, generamos una subclase llamada "portafolio" que permite la interacción de más de un activo y replica los métodos del entorno de un solo activo. Una vez configurado el entorno, procedimos a establecer un escenario de prueba en el que solo utilizamos 2 activos, 2 señales (subirá o bajará el precio) y 3 estados. Los estados se detallan a continuación:

Tabla 4 – Estados del ambiente de prueba

Estado / Inversión	**Activo 1**	**Activo 2**
Estado 1	0%	0%
Estado 2	100%	0%
Estado 3	0%	100%

De esta forma evaluamos si el algoritmo era capaz de aprender la política perfecta para luego agregar los otros 3 activos, aumentar el número de estados y cambiar la señal a una señal de mercado. Para el escenario de prueba el espacio de estado/acción tiene solo 16 opciones, lo que facilita al agente aprender y encontrar la política óptima. Para ello, evaluamos dos tipos de agentes: SARSA y Q-learning. Los resultados obtenidos indican que el agente Q-learning aprende de manera más efectiva que SARSA, por lo que optamos por seleccionarlo para desarrollar el modelo que aborda el problema real.

Figura 1 – Comparación recompensa SARSA vs Q-Learning



Basado en esto se entrenó el agente de Q-learning para ver si podía encontrar la política perfecta que se define a continuación y se muestra el tiempo en que demoró en

aprenderla. Como se mencionó, la señal utilizada en el entorno de prueba es que en la próxima unidad de tiempo el activo va a subir o bajar (una señal que no se tiene en la vida real).

Tabla 5 – Descripción estados, señales y política perfecta

	Señal BTC	Señal ETH	Posicón en BTC	Posición en ETH	Politica perfecta	Justificación
Estado 1	Baja	Baja	0%	0%	Vender todo	Todo va a bajar
Estado 2	Baja	Baja	0%	100%	Vender todo	Todo va a bajar
Estado 3	Baja	Baja	100%	0%	Vender todo	Todo va a bajar
Estado 4	Baja	Sube	0%	0%	Comprar ETH	ETH va a subir y BTC a bajar
Estado 5	Sube	Baja	0%	0%	Comprar BTC	BTC va a subir y ETH a bajar
Estado 6	Baja	Baja	100%	100%	No ocurre	No puede tener 100% del portafolio en dos activos
Estado 7	Baja	Sube	0%	100%	Comprar ETH	ETH va a subir y BTC a bajar
Estado 8	Baja	Sube	100%	0%	Comprar ETH	ETH va a subir y BTC a bajar
Estado 9	Sube	Baja	0%	100%	Comprar BTC	BTC va a subir y ETH a bajar
Estado 10	Sube	Baja	100%	0%	Comprar BTC	BTC va a subir y ETH a bajar
Estado 11	Sube	Sube	0%	0%	Comprar ETH	ETH es más volatil y va a subir más
Estado 12	Baja	Sube	100%	100%	No ocurre	No puede tener 100% del portafolio en dos activos
Estado 13	Sube	Baja	100%	100%	No ocurre	No puede tener 100% del portafolio en dos activos
Estado 14	Sube	Sube	0%	100%	Comprar ETH	ETH es más volatil y va a subir más
Estado 15	Sube	Sube	100%	0%	Comprar ETH	ETH es más volatil y va a subir más
Estado 16	Sube	Sube	100%	100%	No ocurre	No puede tener 100% del portafolio en dos activos

Efectivamente el agente aprendió en pocos episodios la política perfecta. Casi siempre en 50 episodios y siempre en 100.

Tabla 6 – Política aprendida por el agente en distintos episodios

	Politica perfecta	Episodio 1	Episodio 10	Episodio 20	Episodio 30	Episodio 40	Episodio 50	Episodio 100
Estado 1	Vender todo	Vender todo	Vender todo	Vender todo	Vender todo	Vender todo	Vender todo	Vender todo
Estado 2	Vender todo	Vender todo	Vender todo	Vender todo	Vender todo	Vender todo	Vender todo	Vender todo
Estado 3	Vender todo	Vender todo	Vender todo	Vender todo	Vender todo	Vender todo	Vender todo	Vender todo
Estado 4	Comprar ETH	Vender todo	Vender todo	Vender todo	Comprar ETH	Comprar ETH	Comprar ETH	Comprar ETH
Estado 5	Comprar BTC	Vender todo	Vender todo	Comprar ETH	Comprar ETH	Comprar BTC	Comprar BTC	Comprar BTC
Estado 6	No ocurre							
Estado 7	Comprar ETH	Vender todo	Comprar ETH	Comprar ETH	Comprar ETH	Comprar ETH	Comprar BTC	Comprar ETH
Estado 8	Comprar ETH	Vender todo	Comprar ETH	Comprar ETH	Comprar ETH	Comprar ETH	Comprar ETH	Comprar ETH
Estado 9	Comprar BTC	Vender todo	Vender todo	Vender todo	Comprar BTC	Comprar BTC	Comprar BTC	Comprar BTC
Estado 10	Comprar BTC	Vender todo	Comprar ETH	Comprar BTC	Comprar ETH	Comprar BTC	Comprar BTC	Comprar BTC
Estado 11	Comprar ETH	Vender todo	Comprar ETH	Comprar BTC	Vender todo	Comprar BTC	Comprar ETH	Comprar ETH
Estado 12	No ocurre							
Estado 13	No ocurre							
Estado 14	Comprar ETH	Vender todo	Comprar BTC	Comprar BTC	Comprar ETH	Comprar BTC	Comprar ETH	Comprar ETH
Estado 15	Comprar ETH	Vender todo	Vender todo	Comprar BTC	Comprar ETH	Comprar ETH	Comprar ETH	Comprar ETH
Estado 16	No ocurre							
Efectividad politica		25%	50%	50%	75%	83%	92%	100%

Una vez probado que el agente y el ambiente funcionan se escaló el modelo al portafolio esperado y con señales de mercado reales. Se utilizaron los cinco activos identificados anteriormente y como señales se definieron los siguientes indicadores, que son comúnmente utilizados por los negociadores de intradía:

- Precio de apertura
- Precio de cierre
- Precio máximo alcanzado en la unidad de tiempo
- Precio mínimo alcanzado en la unidad de tiempo
- Volumen tranzado
- Promedio simple de 5 unidades de tiempo
- Promedio simple de 8 unidades de tiempo
- Promedio simple de 12 unidades de tiempo

Respecto al ambiente, se establecieron el mismo número de acciones, pero en el caso de los estados aumentaron significativamente pasando solo de 2 por variable a 10 por variable. Esto hace referencia a que las señales se discretizaron inicialmente en dos estados, [0%, 100%] y pasaron a [-100%, -80%, -60%, -40%, -20%, 20%, 40%, 60%, 80% 100%] Esto creo un espacio de estado de 3,200,000 observaciones por 4 distintas acciones. Con estos parámetros, el modelo se entrenó mediante la exploración de distintos hiper parámetros para encontrar el mejor resultado de recompensas. Los parámetros para modificar son: la tasa de descuento, tasa de exploración vs explotación, tasa de aprendizaje número de episodios (cuantas veces el agente a recorrer toda la data). Primero, la tasa de descuento (Gamma) ajusta la importancia de las recompensas futuras en las decisiones del agente. Un valor cercano a 1 indica mayor consideración a largo plazo, mientras que un valor cercano a 0 prioriza las recompensas inmediatas. Segundo, la tasa de aprendizaje (alpha) es un parámetro que controla la magnitud de cuanto se actualiza el valor de una acción en cada actualización. Un valor alto de alpha implica

ajustes más drásticos, mientras que un valor bajo favorece ajustes más graduales y estables. Finalmente, enfriamiento hace referencia a un algoritmo que a medida que pasa el tiempo va reduciendo hasta un número objetivo la tasa de aprendizaje y/o la tasa de exploración/explotación.

El conjunto de datos utilizados correspondió a todas las unidades de tiempo de una hora entre 2021-1 y 2023-06 de los cinco activos seleccionados. Esto representaba alrededor de 22,000 distintos escenarios por activo que se recorrían en un solo episodio.

A continuación, el resultado de estos entrenamientos.

Tabla 7 – Resultado iteración de hiper parámetros

	Prueba				
Categoría	**1**	**2**	**3**	**4**	**5**
Tasa de descuento (gamma)	1	1	0.9	0.8	0.8
Tasa de exploración/explotación (epsilon)	1	0.7	0.5	0.2	Política de enfriamiento que cae 3% por episodio e inicia en 0.8
Tasa de aprendizaje (alpha)	0.1	0.2	0.3	0.4	Política de enfriamiento que cae a 0.1 durante el episodio
Numero de episodios	50	100	150	200	200
Recompensa promedio	**-4**	**3**	**9**	**16**	**59**
Recompensa mínima	-8	-0.6	0.56	0.3	0.5
Recompensa máxima	2.5	6.4	15	28	77

Estos resultados evidencian que el agente efectivamente adquiere una política positiva, al obtener recompensas positivas significativas. En consecuencia, se llevó a cabo una prueba de backtesting, consistente en evaluar el desempeño del agente utilizando datos históricos que no fueron empleados durante el proceso de entrenamiento.

En el estudio, se utilizaron dos muestras de dos mil unidades de tiempo, cada una correspondiente a una hora, en los datos históricos de los cinco activos seleccionados, incluido Bitcoin. La primera muestra abarcó datos desde junio de 2020, caracterizada por una tendencia alcista del mercado, coincidiendo con el periodo en el que Bitcoin experimentó un notable aumento de precios. La segunda muestra abarcó datos desde junio de 2023, evidenciando una tendencia bajista, asociada a un período en el que Bitcoin, junto con otros criptoactivos, experimentó volatilidad y declive en sus valores debido a noticias negativas en el mercado como el colapso de FTX y las altas tasas de interés del periodo.

Para construir el escenario alcista y bajista, se consideró la dirección predominante de la tendencia del mercado en cada periodo, reflejando así las dinámicas específicas de los criptoactivos. La muestra alcista iniciada en junio de 2020 se seleccionó debido al observado comportamiento ascendente de los precios, mientras que la muestra bajista iniciada en junio de 2023 se eligió por la tendencia descendente característica de ese momento. Estas elecciones se basaron en la relevancia de capturar condiciones representativas del mercado de criptoactivos para evaluar la capacidad del modelo en diferentes contextos y basadas en el sentimiento de la comunidad en general.

Es esencial destacar que, al evaluar el desempeño del modelo con una tasa de exploración igual a cero y un capital de USD$ 1,000 en ambos casos, se compararon los rendimientos obtenidos con respecto a una estrategia de asignación pasiva, teniendo en cuenta la peculiar volatilidad y naturaleza estocástica del mercado de criptoactivos. Además, se observó que el modelo ajustaba sus valores Q y su política de manera más lenta en comparación con el escenario de entrenamiento, un fenómeno coherente con la variabilidad inherente a los precios de criptoactivos como Bitcoin. Los resultados de la prueba se observan a continuación:

Figura 2. Valor del portafolio en escenario alcista

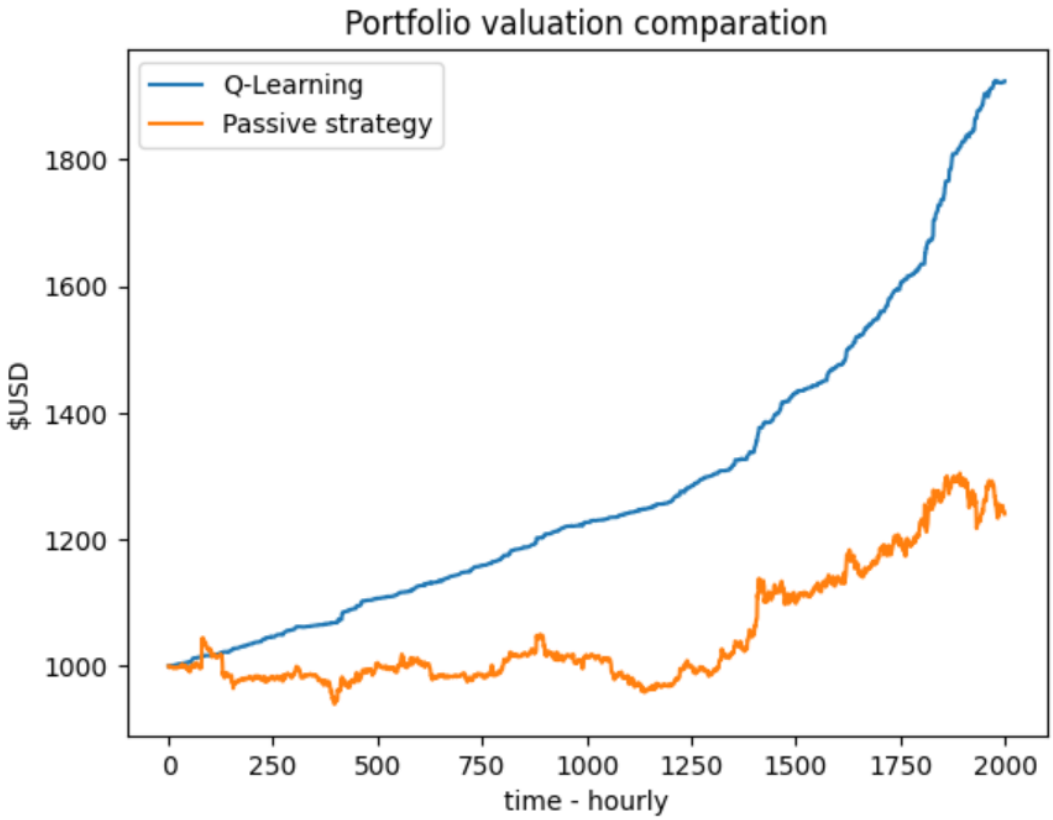
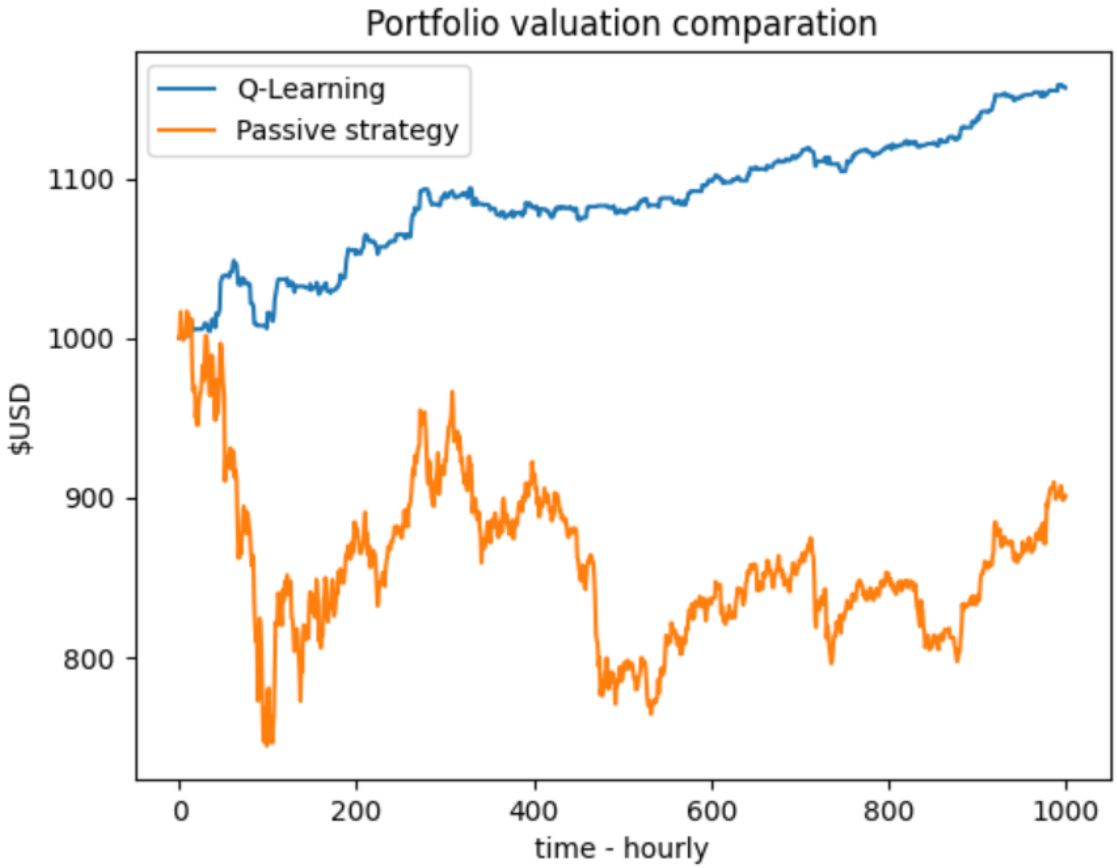


Figura 3. Valor del portafolio en escenario bajista



Como se puede observar en las gráficas, no hay duda de que el agente aprende y supera la política pasiva en el escenario alcista y en el escenario bajista. Esto demuestra que el aprendizaje por refuerzo sí es capaz de detectar señales en las distintas variables del estado que el agente tiene a su disposición. Sin embargo, es necesario expandir las pruebas realizadas y abarcar un testeo de mayor profundidad para asegurar que los resultados son permanentes en más escenarios.

Continuando con la metodología y para ilustrar la política de inversión del agente, se seleccionó el cambio en el precio de los activos y su magnitud como la variable principal para describir un estado. Esto significa que a cada activo se le asignó una magnitud, que varía entre uno y tres, representada por flechas. Las flechas hacia arriba indican una tendencia alcista en el precio del activo, mientras que las flechas hacia abajo señalan una tendencia bajista. Esta metodología permitió identificar los 10 estados más relevantes en los cuales el agente obtuvo mayor recompensa al invertir.

Tabla 8 – Ranking 10 estados de mayor recompensa

Estado	**Ranking**
BTC ↑, ETH ↑, BNB ↑↑↑, TRX ↑, DOT ↓	1
BTC ↑, ETH ↑, BNB ↓, TRX ↑, DOT ↑↑↑	2
BTC ↑, ETH ↑, BNB ↑, TRX ↓, DOT ↑↑	3
BTC ↑↑, ETH ↑, BNB ↓, TRX ↑, DOT ↑	4
BTC ↑, ETH ↑↑, BNB ↓, TRX ↑, DOT ↑	5
BTC ↑, ETH ↑, BNB ↑, TRX ↑↑, DOT ↓	6
BTC ↓, ETH ↑, BNB ↑, TRX ↑↑, DOT ↑	7
BTC ↑↑, ETH ↑, BNB ↑, TRX ↓, DOT ↑	8
BTC ↑, ETH ↑, BNB ↓, TRX ↑, DOT ↑	9
BTC ↑↑, ETH ↑, BNB ↑, TRX ↑, DOT ↓	10

Además, se elaboró una tabla para facilitar la comprensión de las decisiones del agente en cada estado específico. Por ejemplo, en esta tabla se observa que, en periodos de alta volatilidad, caracterizados por cambios pronunciados en la tendencia de precios

(indicados por 3 flechas hacia arriba o hacia abajo), el algoritmo tiende a abstenerse de invertir, casos que se identifican con color naranja. Asimismo, se destaca que una tendencia bajista significativa en el precio de Bitcoin puede ser interpretada por el algoritmo como una oportunidad de inversión favorable, como se muestra en la tabla, donde esta situación se marca en verde como se evidencia a continuación:

Tabla 9 – Política, estado y acciones

Politica						Estado y acción		
No invierte	Invierte BTC	Invierte ETH	Invierte BNB	Invierte TRX	Invierte DOT	Estado	Que hace	Activo inversion
0.0%	0.0%	99.8%	0.0%	0.0%	0.0%	BTC ↓↓↓,ETH ↓↓,BNB ↑,TRX ↑,DOT ↓	Invierte	ETH
0.0%	0.0%	99.8%	0.0%	0.0%	0.0%	BTC ↓↓↓,ETH ↓↓,BNB ↓,TRX ↓↓,DOT ↓↓	Invierte	ETH
0.0%	0.0%	99.8%	0.0%	0.0%	0.0%	BTC ↓↓↓,ETH ↓↓,BNB ↓,TRX ↓,DOT ↑↑↑	Invierte	ETH
0.0%	99.8%	0.0%	0.0%	0.0%	0.0%	BTC ↓↓↓,ETH ↓↓,BNB ↑↑↑,TRX ↓,DOT ↓	Invierte	BTC
0.0%	0.0%	0.0%	0.0%	99.8%	0.0%	BTC ↓↓↓,ETH ↓,BNB ↓↓,TRX ↓,DOT ↓	Invierte	TRX
0.0%	0.0%	0.0%	99.8%	0.0%	0.0%	BTC ↓↓↓,ETH ↓,BNB ↓↓,TRX ↑,DOT ↓	Invierte	BNB
0.0%	0.0%	0.0%	99.8%	0.0%	0.0%	BTC ↓↓↓,ETH ↓,BNB ↓,TRX ↓,DOT ↓	Invierte	BNB
0.0%	0.0%	99.8%	0.0%	0.0%	0.0%	BTC ↓↓↓,ETH ↓,BNB ↓,TRX ↓,DOT ↑	Invierte	ETH
0.0%	99.8%	0.0%	0.0%	0.0%	0.0%	BTC ↓↓↓,ETH ↓,BNB ↓,TRX ↓,DOT ↓↓	Invierte	BTC
0.0%	0.0%	0.0%	99.8%	0.0%	0.0%	BTC ↓↓↓,ETH ↓,BNB ↑,TRX ↓,DOT ↑↑↑	Invierte	BNB
0.0%	0.0%	0.0%	0.0%	0.0%	99.8%	BTC ↓↓↓,ETH ↓,BNB ↑↑,TRX ↓,DOT ↓	Invierte	DOT
0.0%	0.0%	0.0%	99.8%	0.0%	0.0%	BTC ↓↓↓,ETH ↓,BNB ↓↓,TRX ↓,DOT ↑	Invierte	BNB
0.0%	0.0%	99.8%	0.0%	0.0%	0.0%	BTC ↓↓↓,ETH ↓,BNB ↑,TRX ↓,DOT ↑	Invierte	ETH
0.0%	0.0%	0.0%	0.0%	99.8%	0.0%	BTC ↓↓↓,ETH ↓,BNB ↑↑↑,TRX ↑,DOT ↑	Invierte	TRX
0.0%	0.0%	0.0%	0.0%	99.8%	0.0%	BTC ↓↓↓,ETH ↓,BNB ↓↓,TRX ↓↓,DOT ↓↓	Invierte	TRX
0.0%	99.8%	0.0%	0.0%	0.0%	0.0%	BTC ↓↓↓,ETH ↑,BNB ↓↓,TRX ↓,DOT ↑	Invierte	BTC
0.0%	99.8%	0.0%	0.0%	0.0%	0.0%	BTC ↓↓↓,ETH ↑,BNB ↓,TRX ↓,DOT ↓	Invierte	BTC
0.0%	0.0%	0.0%	99.8%	0.0%	0.0%	BTC ↓↓↓,ETH ↑,BNB ↓,TRX ↓↓,DOT ↑↑	Invierte	BNB
0.0%	99.8%	0.0%	0.0%	0.0%	0.0%	BTC ↓↓↓,ETH ↑,BNB ↓,TRX ↑,DOT ↑	Invierte	BTC
0.0%	0.0%	0.0%	0.0%	0.0%	99.8%	BTC ↓↓↓,ETH ↑,BNB ↑,TRX ↓↓,DOT ↑	Invierte	DOT
99.8%	0.0%	0.0%	0.0%	0.0%	0.0%	BTC ↓↓↓,ETH ↓↓↓,BNB ↓↓,TRX ↑↑↑,DOT ↑↑↑	No invierte	
99.8%	0.0%	0.0%	0.0%	0.0%	0.0%	BTC ↓↓↓,ETH ↓↓↓,BNB ↑↑↑,TRX ↑↑,DOT ↓↓	No invierte	
99.8%	0.0%	0.0%	0.0%	0.0%	0.0%	BTC ↓↓↓,ETH ↓↓,BNB ↑↑↑,TRX ↑,DOT ↓	No invierte	
99.8%	0.0%	0.0%	0.0%	0.0%	0.0%	BTC ↓↓↓,ETH ↓,BNB ↓↓,TRX ↓↓,DOT ↓	No invierte	
99.8%	0.0%	0.0%	0.0%	0.0%	0.0%	BTC ↓↓↓,ETH ↓,BNB ↓↓,TRX ↑↑↑,DOT ↓	No invierte	
99.8%	0.0%	0.0%	0.0%	0.0%	0.0%	BTC ↓↓↓,ETH ↓,BNB ↑,TRX ↓↓,DOT ↑	No invierte	
99.8%	0.0%	0.0%	0.0%	0.0%	0.0%	BTC ↓↓,ETH ↓↓,BNB ↓↓,TRX ↓↓,DOT ↓↓↓	No invierte	
99.8%	0.0%	0.0%	0.0%	0.0%	0.0%	BTC ↓↓,ETH ↓↓,BNB ↓↓,TRX ↓↓,DOT ↓↓	No invierte	
99.8%	0.0%	0.0%	0.0%	0.0%	0.0%	BTC ↓↓,ETH ↓↓,BNB ↓↓,TRX ↓↓↓,DOT ↓↓	No invierte	
99.8%	0.0%	0.0%	0.0%	0.0%	0.0%	BTC ↓↓,ETH ↓↓,BNB ↓↓,TRX ↓↓,DOT ↓↓	No invierte	
99.8%	0.0%	0.0%	0.0%	0.0%	0.0%	BTC ↓↓,ETH ↓↓↓,BNB ↓↓,TRX ↓↓,DOT ↓↓↓	No invierte	

Finalmente, también se hizo un análisis para determinar posibles variables de decisión para el agente y así tratar de comprender cuál es el raciocinio detrás de la política descubierta por el agente. Se determinó si había algún tipo de relación entre las señales y el comportamiento del agente donde se identificó la relevancia de la diferencia absoluta de los promedios de 5, 8 y 12 unidades de tiempo de todos los activos. En esta se observa

que cuando esta diferencia se acerca al 0.6 el agente ve más probable invertir en distintos activos en especial en Bitcoin.

Tabla 10 – Probabilidad de inversión dada la diferencia absoluta de promedio simples

Señal más relevante*	BNB	BTC	DOT	ETH	TRX
5	0.0%	0.0%	0.0%	0.0%	0.0%
4.8	0.0%	0.0%	0.0%	0.0%	0.0%
4.6	0.0%	0.0%	0.0%	0.0%	0.0%
4.3	0.0%	0.1%	0.0%	0.1%	0.0%
4.1	0.1%	0.1%	0.0%	0.1%	0.1%
3.9	0.1%	0.1%	0.1%	0.1%	0.1%
3.7	0.1%	0.2%	0.1%	0.1%	0.1%
3.4	0.2%	0.3%	0.2%	0.2%	0.2%
3.2	0.3%	0.3%	0.2%	0.3%	0.2%
3	0.3%	0.4%	0.3%	0.3%	0.3%
2.8	0.4%	0.4%	0.3%	0.5%	0.3%
2.6	0.5%	0.5%	0.4%	0.4%	0.4%
2.3	0.5%	0.6%	0.4%	0.6%	0.4%
2.1	0.6%	0.8%	0.4%	0.7%	0.6%
1.9	0.7%	0.8%	0.5%	0.8%	0.5%
1.7	0.7%	0.7%	0.5%	0.8%	0.6%
1.4	0.7%	0.8%	0.6%	0.8%	0.7%
1.2	0.7%	0.9%	0.7%	0.8%	0.7%
1	0.7%	0.9%	0.6%	0.9%	0.8%
0.8	0.8%	0.9%	0.8%	1.0%	0.8%
0.6	0.9%	1.1%	0.7%	1.0%	0.8%
0.3	0.7%	1.0%	0.6%	0.9%	0.7%
0.1	0.7%	0.9%	0.7%	0.8%	0.7%
-0.1	0.8%	0.7%	0.6%	0.9%	0.7%
-0.3	0.7%	0.9%	0.6%	0.9%	0.6%
-0.6	0.6%	0.7%	0.5%	0.6%	0.6%
-0.8	0.5%	0.6%	0.4%	0.6%	0.5%
-1	0.5%	0.6%	0.4%	0.5%	0.4%
-1.2	0.4%	0.5%	0.4%	0.5%	0.3%
-1.4	0.3%	0.4%	0.2%	0.4%	0.3%
-1.7	0.2%	0.3%	0.2%	0.3%	0.2%
-1.9	0.2%	0.2%	0.1%	0.2%	0.2%
-2.1	0.1%	0.2%	0.1%	0.2%	0.2%
-2.3	0.1%	0.1%	0.1%	0.1%	0.1%
-2.6	0.1%	0.1%	0.1%	0.1%	0.1%
-2.8	0.0%	0.1%	0.1%	0.0%	0.0%
-3	0.0%	0.0%	0.0%	0.0%	0.0%
-3.2	0.0%	0.0%	0.0%	0.0%	0.0%
-3.4	0.0%	0.0%	0.0%	0.0%	0.0%
-3.7	0.0%	0.0%	0.0%	0.0%	0.0%

*Differencia absoluta de promedios simples (5,8,12)

Capítulo 7
CONCLUSIONES Y RECOMENDACIONES

La administración de portafolios de criptoactivos utilizando técnicas de aprendizaje por refuerzo representa un avance significativo en la aplicación de inteligencia artificial en el ámbito financiero. Sin embargo, hay aspectos que pueden ser explorados y mejorados en futuros desarrollos.

Una conclusión clave es la capacidad del modelo para adaptarse a las fluctuantes condiciones del mercado, lo que resulta esencial para manejar el riesgo y maximizar los retornos en un ámbito tan impredecible como se pudo ver en las pruebas de mercado alcista y bajista. Al compararlo con estrategias de gestión de portafolios tradicionales, el modelo de aprendizaje por refuerzo muestra un desempeño superior, destacando su potencial para transformar las prácticas de inversión en criptoactivos.

Una de las áreas clave para el desarrollo futuro es la implementación de técnicas de aprendizaje profundo para estimar los valores Q en entornos con espacios de estado/acción de grandes dimensiones. En el contexto del mercado de criptoactivos, que se caracteriza por su complejidad y dinamismo, los modelos tradicionales de Q-learning pueden enfrentar limitaciones cuando se trata de manejar una gran cantidad de variables y relaciones no lineales. Aquí, el aprendizaje profundo puede jugar un papel crucial, ofreciendo una capacidad superior para procesar y aprender de vastos conjuntos de datos, lo que permite identificar patrones y tendencias que de otra forma serían difíciles de captar.

Además, para aumentar la precisión y efectividad del modelo, sería beneficioso transformar las señales utilizadas de discretas a continuas. Esto permitiría al modelo capturar matices más finos en los datos del mercado y responder con mayor precisión a cambios sutiles en las condiciones del mercado. Por ejemplo, en lugar de usar señales

simples como "comprar" o "vender", se podrían emplear señales basadas en probabilidades o rangos de confianza, lo que proporcionaría una comprensión más matizada de las tendencias del mercado.

Otra área de mejora sería la inclusión y experimentación con una variedad más amplia de señales de trading. En lugar de limitarse a indicadores convencionales como los promedios móviles o el volumen de transacciones, se podrían integrar señales más avanzadas como el análisis de sentimientos basado en noticias financieras, indicadores de tendencia, o incluso algoritmos de predicción de tendencias basados en aprendizaje de máquina. Esto no solo mejoraría la capacidad del modelo para realizar predicciones precisas, sino que también permitiría una adaptación más efectiva a las condiciones cambiantes del mercado.

Finalmente, es crucial que estos desarrollos se realicen considerando las implicaciones éticas y regulatorias, especialmente en un campo tan novedoso y en rápida evolución como el de los criptoactivos. A medida que la tecnología avanza, también lo hacen las expectativas y normativas en torno a la transparencia, la seguridad y la responsabilidad en la gestión financiera.

En resumen, mientras el trabajo actual proporciona una base sólida, la integración de aprendizaje profundo para la estimación de valores Q, la transformación de señales de discretas a continuas y la exploración de nuevas señales de trading representan pasos importantes hacia el perfeccionamiento de modelos de gestión de portafolios en el ámbito de los criptoactivos.

REFERENCIAS

1. Al-Aradi, A., & Jaimungal, S. (2019). Active and Passive Portfolio Management with Latent Factors. Recuperado de http://arxiv.org/abs/1903.06928v1
2. Bertoluzzo, F., & Corazza, M. (2012). Testing Different Reinforcement Learning Configurations for Financial Trading: Introduction and Applications. Recuperado de https://www.researchgate.net/publication/257744845_Testing_Different_Reinforcement_Learning_Configurations_for_Financial_Trading_Introduction_and_Applications
3. Brockman, G., et al. (2016). OpenAI Gym. arXiv preprint arXiv:1606.01540
4. Bufalo, D., Bufalo, M., Cesarone, F., & Orlando, G. (2022). Straightening skewed markets with an index tracking optimizationless portfolio. Recuperado de http://arxiv.org/abs/2203.13766v1
5. Çela, E., Hafner, S., Mestel, R., & Pferschy, U. (2022). Integrating multiple sources of ordinal information in portfolio optimization. Recuperado de http://arxiv.org/abs/2211.00420v2
6. Charles-Cadogan, G. (2012). Active Portfolio Management, Positive Jensen-Jarrow Alpha, and Zero Sets of CAPM. Recuperado de http://arxiv.org/abs/1206.4562v1
7. Chakraborty, S., & Joseph, A. (2017). Machine Learning at Central Banks. Bank of England Working Paper No. 674
8. Chakraborty, S. (2019). Capturing Financial markets to apply Deep Reinforcement Learning. Retrieved from arXiv:1907.04373v3
9. Elendner, H., Trimborn, S., Ong, B., & Lee, T. M. (2018). The Cross-Section of Crypto-Currencies as Financial Assets: An Overview. Investment Management and Financial Innovations, 15(4), 33-49.

10. Fang, F., Chung, W., Ventre, C., Basios, M., Kanthan, L., Li, L., & Wu, F. (2020). Ascertaining Price Formation in Cryptocurrency Markets with DeepLearning. arXiv.
11. Filos, A. (2019). Reinforcement Learning for Portfolio Management. Recuperado de http://arxiv.org/abs/1909.09571v1
12. Gao, Z., Gao, Y., Hu, Y., Jiang, Z., & Su, J. (2020). Application of Deep Q-Network in Portfolio Management. Recuperado de http://arxiv.org/abs/2003.06365v1
13. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT press.
14. Gruszka, J., & Szwabiński, J. (2023). Portfolio Optimisation via the Heston Model Calibrated to Real Asset Data. Recuperado de http://arxiv.org/abs/2302.01816v1
15. Huang, G., Zhou, X., & Song, Q. (2020). Deep reinforcement learning for portfolio management**.** Recuperado de arXiv:2012.13773v7
16. Jiang, Z., & Liang, J. (2016). Cryptocurrency Portfolio Management with Deep Reinforcement Learning. Recuperado de http://arxiv.org/abs/1612.01277v5
17. Jiang, Z., & Liang, J. (2018). Financial Trading as a Game: A Deep Reinforcement Learning Approach. Recuperado de http://arxiv.org/abs/1807.02787v1
18. Jiang, Z., Xu, D., & Liang, J. (2017). A Deep Reinforcement Learning Framework for the Financial Portfolio Management Problem. arXiv preprint arXiv:1706.10059.
19. Jing, R., & Rocha, L. E. C. (2023). A network-based strategy of price correlations for optimal cryptocurrency portfolios. Recuperado de http://arxiv.org/abs/2304.02362v1

20. Li, X., Wu, P., Zou, C., & Li, Q. (2022). Hierarchical Deep Reinforcement Learning for VWAP Strategy Optimization. arXiv. http://arxiv.org/abs/2212.14670v1
21. Li, Y., & Hoi, S. C. (2014). Online Portfolio Selection: A Survey. ACM Computing Surveys (CSUR), 46(3), 1-36.
22. Li, Y., Wang, J., & Cao, Y. (2019). A General Framework on Enhancing Portfolio Management with Reinforcement Learning. arXiv preprint arXiv:1911.11880.
23. Liu, Y., & Tsyvinski, A. (2018). Risks and Returns of Cryptocurrency. The Review of Financial Studies, Oxford Academic.
24. Liew, J., & Budavári, T. (2013). Adaptive Portfolio Management with Reinforcement Learning. In Proceedings of the ... International Conference on Machine Learning. Workshop on Machine Learning for Global Risk Management.
25. Mnih, V., et al. (2015). Human-level control through deep reinforcement learning. Nature, 518(7540), 529-533.
26. Moody, J., & Saffell, M. (2001). Learning to trade via direct reinforcement. IEEE transactions on neural Networks, 12(4), 875-889.
27. Murphy, K. P. (2012). Machine Learning: A Probabilistic Perspective. MIT press.
28. Noguer i Alonso, M., & Srivastava, S. (2020). Deep Reinforcement Learning for Asset Allocation in US Equities. Recuperado de http://arxiv.org/abs/2010.04404v1
29. Pigorsch, U., & Schäfer, S. (2021). High-Dimensional Stock Portfolio Trading with Deep Reinforcement Learning. Recuperado de http://arxiv.org/abs/2112.04755v1
30. Saeidi, S. A., Fallah, F., Barmaki, S., & Farbeh, H. (2022). A Novel Neuromorphic Processors Realization of Spiking Deep Reinforcement Learning for Portfolio Management. Recuperado de http://arxiv.org/abs/2203.14159v1

31. Sharpe, W. F. (1994). The Sharpe Ratio. Journal of Portfolio Management, 21(1), 49-58.
32. Spooner, T., Fearnley John, Savani Rahul (2017). Market Making via Reinforcement Learning. Recuperado de http://arxiv.org/abs/1804.04216v1
33. Shin, W., Bu, S.-J., & Cho, S.-B. (2019). Automatic Financial Trading Agent for Low-risk Portfolio Management using Deep Reinforcement Learning. Recuperado de http://arxiv.org/abs/1909.03278v1
34. Silver, D., et al. (2016). Mastering the game of Go with deep neural networks and tree search. Nature, 529(7587), 484-489.
35. Steven Gordon, Zhi Li, John Marthinsen. (2023). A Deep Analysis of the Economics and Finance Research on Cryptocurrencies. ScienceDirect.
36. Sutton, R. S., & Barto, A. G. (2018). Reinforcement Learning: An Introduction. MIT press.
37. Van Heerden, N. A., Cabral, J. B., & Luczywo, N. (2021). Evaluation of the Importance of Criteria for the Selection of Cryptocurrencies. arXiv.
38. Velay, M., Doan, B.-L., Rimmel, A., Popineau, F., & Daniel, F. (2023). Benchmarking Robustness of Deep Reinforcement Learning approaches to Online Portfolio Management. Recuperado de: arXiv:2306.10950v1
39. Wang, R., Wei, H., An, B., Feng, Z., & Yao, J. (2020). Deep Stock Trading: A Hierarchical Reinforcement Learning Framework for Portfolio Optimization and Order Execution. https://arxiv.org/abs/2012.12620v2
40. Watkins, C. J., & Dayan, P. (1992). Q-learning. Machine learning, 8(3-4), 279-292.
41. Härdle, W. K., Harvey, C. R., & Reule, R. C. G. (2020). Editorial: Understanding Cryptocurrencies. Retrieved from arXiv:2007.14702v1
42. Zaghloul, E., Li, T., Mutka, M., & Ren, J. (2019). Bitcoin and Blockchain: Security and Privacy. Retrieved from arXiv:1904.11435v1