



**UNIVERSIDAD ESTATAL PENÍNSULA
DE SANTA ELENA
FACULTAD DE SISTEMAS Y
TELECOMUNICACIONES**

TITULO DEL TRABAJO DE TITULACIÓN

Desarrollo de Estrategia de trading algorítmico usando aprendizaje por
refuerzo

AUTOR

Chancay González Anthony Joel

EXAMEN COMPLEXIVO

Previo a la obtención del grado académico en
INGENIERO EN TECNOLOGÍAS DE LA INFORMACIÓN

TUTOR

Ing. Coronel Suárez Iván Alberto, Msia.

Santa Elena, Ecuador

Año 2025



**UNIVERSIDAD ESTATAL PENÍNSULA
DE SANTA ELENA
FACULTAD DE SISTEMAS Y TELECOMUNICACIONES**

TRIBUNAL DE SUSTENTACIÓN

Ing. José Sánchez A. Msc.
DIRECTOR DE LA CARRERA

Ing. Iván Coronel S. Msia.
TUTOR

Ing. Carlos Castillo Y. Mgt.
DOCENTE ESPECIALISTA

Ing. Marjorie Coronel S. Mgti.
DOCENTE GUÍA UIC



**UNIVERSIDAD ESTATAL PENÍNSULA
DE SANTA ELENA
FACULTAD DE SISTEMAS Y TELECOMUNICACIONES**

CERTIFICACIÓN

Certifico que luego de haber dirigido científica y técnicamente el desarrollo y estructura final del trabajo, este cumple y se ajusta a los estándares académicos, razón por el cual apruebo en todas sus partes el presente trabajo de titulación que fue realizado en su totalidad por **Chancay González Anthony Joel**, como requerimiento para la obtención del título de Ingeniero en Tecnologías de la Información.

La Libertad, a los 20 días del mes de Junio del año 2025

TUTOR



ING. IVÁN CORONEL, MSIA.



**UNIVERSIDAD ESTATAL PENÍNSULA
DE SANTA ELENA
FACULTAD DE SISTEMAS Y TELECOMUNICACIONES**

DECLARACIÓN DE RESPONSABILIDAD

Yo, Chancay González Anthony Joel

DECLARO QUE:

El trabajo de Titulación, desarrollo de estrategia de trading algorítmico usando aprendizaje por refuerzo previo a la obtención del título en Ingeniero en Tecnologías de la Información, ha sido desarrollado respetando derechos intelectuales de terceros conforme las citas que constan en el documento, cuyas fuentes se incorporan en las referencias o bibliografías. Consecuentemente este trabajo es de mi total autoría.

En virtud de esta declaración, me responsabilizo del contenido, veracidad y alcance del Trabajo de Titulación referido.

La Libertad, a los 20 días del mes de Junio del año 2025

EL AUTOR

A handwritten signature in blue ink that reads "Anthony Chancay".

Anthony Joel Chancay González




**UNIVERSIDAD ESTATAL PENÍNSULA
DE SANTA ELENA**

FACULTAD DE SISTEMAS Y TELECOMUNICACIONES

CERTIFICACIÓN DE ANTIPLAGIO

Certifico que después de revisar el documento final del trabajo de titulación denominado **Desarrollo de Estrategia de trading algorítmico usando aprendizaje por refuerzo**, presentado por el estudiante, **Chancay González Anthony Joel** fue enviado al Sistema Antiplagio, presentando un porcentaje de similitud correspondiente al 4%, por lo que se aprueba el trabajo para que continúe con el proceso de titulación.

 CERTIFICADO DE ANÁLISIS
magister

**CHANCAY GONZALEZ ANTHONY
JOEL (3)**

4%
Textos sospechosos



< 1% Similitudes
0% similitudes entre comillas
0% entre las fuentes mencionadas

3% Idiomas no reconocidos

12% Textos potencialmente generados por IA (ignorado)

Nombre del documento: CHANCAY GONZALEZ ANTHONY JOEL (3).pdf ID del documento: 4651de6982bbccce9cf643ae5976e253db4189c5 Tamaño del documento original: 815,19 kB	Depositante: IVAN ALBERTO CORONEL SUAREZ Fecha de depósito: 21/6/2025 Tipo de carga: interface fecha de fin de análisis: 21/6/2025	Número de palabras: 10.581 Número de caracteres: 70.335
---	---	--

Ubicación de las similitudes en el documento:



Firmado electrónicamente por:
**IVAN ALBERTO
CORONEL SUAREZ**
Validar únicamente con FirmaRC

Ing. Iván Coronel Suárez, Msia

TUTOR



**UNIVERSIDAD ESTATAL PENÍNSULA
DE SANTA ELENA
FACULTAD DE SISTEMAS Y TELECOMUNICACIONES**

AUTORIZACIÓN

Yo, Chancay González Anthony Joel

Autorizo a la Universidad Estatal Península de Santa Elena, para que haga de este trabajo de titulación o parte de él, un documento disponible para su lectura consulta y procesos de investigación, según las normas de la Institución.

Cedo los derechos en línea patrimoniales del presente trabajo de titulación con fines de difusión pública, además apruebo la reproducción dentro de las regulaciones de la Universidad, siempre y cuando esta reproducción no suponga una ganancia económica y se realice respetando mis derechos de autor

Santa Elena, a los 20 días del mes de Junio del año 2025

EL AUTOR

A handwritten signature in blue ink that reads "Anthony Chancay".

Anthony Joel Chancay González

AGRADECIMIENTO

Para empezar, agradezco inmensamente a Dios, quien me sostuvo y fue mi guía a lo largo de esta carrera, acompañándome y bendiciéndome de muchísimas formas en cada momento.

Agradezco a mis padres Rubén y Aracely, quienes me dieron la vida, y sacrificaron mucho para darme la educación y todo lo que pudiera necesitar a lo largo de mi vida.

Agradezco a mis hermanos Jefferson y Michael, por ser mi compañía y apoyarme. A mis hermanos y amigos de la Iglesia, Bryan, Jeremy y a todos gracias por siempre estar.

A cada persona que me apoyó con una palabra, un gesto, un ánimo, muchas gracias.

Anthony Joel, Chancay González

DEDICATORIA

Este trabajo va dedicado a Dios, por Él tengo todo y hacia Él va todo. A mis padres y hermanos, quienes fueron mi fuente de inspiración para ser constante y no desistir, quienes me enseñaron con su ejemplo a esforzarme y perseverar para alcanzar mis metas.

A mi abuelo que está en el cielo, pero siempre creyó en mí y confiaba en que podría cumplir lo que me proponía.

A todos los que creyeron y confiaron en mí, a los que empiezan sus carreras y a quien lee esto, se los dedico como muestra de que el ser constante y sacrificarse da frutos.

Anthony Joel, Chancay González

ÍNDICE GENERAL

TITULO DEL TRABAJO DE TITULACIÓN	I
TRIBUNAL DE SUSTENTACIÓN	II
CERTIFICACIÓN	III
DECLARACIÓN DE RESPONSABILIDAD	IV
DECLARO QUE:	IV
CERTIFICACIÓN DE ANTIPLAGIO	V
AUTORIZACIÓN	VI
AGRADECIMIENTO	VII
DEDICATORIA	VIII
ÍNDICE GENERAL	IX
ÍNDICE DE TABLAS	XIII
ÍNDICE DE FIGURAS	XIV
RESUMEN	XV
ABSTRACT	XV
INTRODUCCIÓN	2
CAPÍTULO 1. FUNDAMENTACIÓN	3
1.1 Antecedentes	3
1.2 Descripción del Proyecto	5
1.3 Objetivos del Proyecto	7
1.4 Justificación del Proyecto	8
1.5 Alcance del Proyecto	9
CAPÍTULO 2. MARCO TEÓRICO Y METODOLOGÍA DEL PROYECTO	10
2.1 Marco Conceptual	10
2.1.1 Aprendizaje por Refuerzo	10

2.1.1.1	Aplicación al Trading	10
2.1.2	Trading Algorítmico	10
2.1.2.1	Backtesting	10
2.1.3	Búsqueda de rejilla	11
2.1.4	Criptomoneda	11
2.1.4.1	Bitcoin (BTC)	11
2.1.4.2	Ethereum (ETH)	11
2.1.4.3	Tether (USDT)	12
2.1.5	Plataforma de intercambio de criptomonedas	12
2.1.5.1	Binance	12
2.1.6	Mercado de Futuros de Criptomonedas	12
2.1.6.1	Operación	13
2.1.6.2	Órdenes de Trading	13
2.1.6.3	Análisis Técnico	14
2.1.7	Métricas de Evaluación	15
2.1.7.1	Ratio de Sharpe	16
2.1.7.2	Drawdown Máximo	16
2.1.7.3	Ratio de ganancias	16
2.1.8	Python.	16
2.1.8.1	Gymnasium	17
2.1.8.2	Pandas.	17
2.1.8.3	Python-binance.	17
2.1.9	Archivos CSV	17

2.2	Marco Teórico	18
2.2.1	Teoría de Ondas de Elliott	18
2.2.1.1	Estructura de las Ondas Principales	18
2.2.2	Teoría de la fractalidad de los Mercados	19
2.3	Metodología del Proyecto	19
2.3.1	Metodología de Investigación	19
2.3.1.1	Contexto de la investigación	19
2.3.1.2	Enfoque de la investigación	20
2.3.1.3	Alcance de la investigación	21
2.3.1.4	Tipo de investigación	21
2.3.1.5	Método de Investigación	21
2.3.1.6	Población y muestra	21
2.3.2	Técnicas e instrumentos de recolección de datos	22
2.3.3	Metodología de desarrollo	22
2.3.3.1	Fase de recolección de datos.	22
2.3.3.2	Fase de optimización de indicadores técnicos.	23
2.3.3.3	Fase de entrenamiento de agente.	23
2.3.3.4	Fase de evaluación de rendimiento.	23
2.3.3.5	Fase de integración a la API de Binance.	23
CAPÍTULO 3.	Propuesta	24
3.1	Fase de recolección de datos	24
3.1.1	Datos de la API de Binance	24
3.1.2	Cálculo de indicadores técnicos	25
3.2	Fase de optimización de indicadores técnicos	26

3.2.1	Búsqueda de combinaciones de parámetros	26
3.2.2	Comparación de resultados obtenidos	27
3.3	Fase de entrenamiento del agente	29
3.3.1	Ambiente virtual de simulación y espacio de observación	30
3.3.2	Espacio de acciones	32
3.3.3	Sistema de recompensas	32
3.3.4	Proceso de entrenamiento	33
3.4	Fase de evaluación del agente	34
3.5	Fase de integración a la API de Binance	36
3.5.1	Arquitectura del sistema de integración	36
CONCLUSIONES		38
RECOMENDACIONES		39
REFERENCIAS		40
ANEXOS		43
Anexo 1: Gráfico de la recompensa media obtenida durante la fase de entrenamiento del agente de aprendizaje por refuerzo.		43
Anexo 2: Comparación de ratio de ganancias y pérdidas entre el backtesting aplicado a la estrategia inicial y el agente entrenado.		43
Anexo 3: Gráfico de dispersión de drawdown máximo y ratio de sharpe entre el backtesting inicial y el agente.		44
Anexo 4: Rentabilidad absoluta y beneficio de la evaluación inicial y final.		44
Anexo 5: Historial de balances durante el backtesting inicial durante 647 operaciones ejecutadas en un año.		45
Anexo 6: Historial de balance durante el backtesting del agente en 3079 operaciones ejecutadas en un año.		45
Anexo 7: Diagrama de flujo del sistema de integración a la API de Binance.		46
Anexo 8: Agente operando en tiempo real en la plataforma de binance		46

ÍNDICE DE TABLAS

Tabla 1 Archivos generados al recolectar datos.	24
Tabla 2 Descripción de datos recolectados desde el api de Binance.	25
Tabla 3 Descripción de datos de indicadores técnicos calculados.	26
Tabla 4 Cambios en los parámetros antes y después de la búsqueda de rejilla.	27
Tabla 5 Descripción de variables internas de observación del agente	31
Tabla 6 Espacio de acciones posibles del agente de aprendizaje por refuerzo	32
Tabla 7 Resultados obtenidos de la estrategia inicial y agente RL	35

ÍNDICE DE FIGURAS

Ilustración 1 Aprendizaje por refuerzo Pröllochs & Feuerriegel, 2016.	6
Ilustración 2 Ondas de Elliot.	18
Ilustración 3 Mercado de futuros de Ethereum en temporalidad 5 minutos	20
Ilustración 4 Gráfico mensual de los datos seleccionados como muestra	22
Ilustración 5 Gráfico comparativo de la rentabilidad obtenido en 3 años	27
Ilustración 6 Distribución de la ratio de operaciones ganadas y perdidas	28
Ilustración 7 Balances máximos, mínimos y finales	28
Ilustración 8 Variación de la ratio de Sharpe y drawdown máximo.	29
Ilustración 9 Recompensas obtenidas en cada episodio	33
Ilustración 10 Ratio de operaciones ganadas en cada episodio	34
Ilustración 11 Esquema del sistema de integración entre el agente de RL y la API de Binance	37

RESUMEN

El objetivo de este proyecto fue desarrollar una estrategia de trading algorítmico mediante el uso de aprendizaje por refuerzo, aplicada al mercado de futuros de Ethereum en la plataforma de intercambios Binance. El método incluyó: recolección de datos históricos de velas de 5 minutos vía API de Binance, optimización de indicadores técnicos mediante grid search, entrenamiento de un agente con datos de 2021-2023, evaluación mediante backtesting en 2023-2024 e integración a la API de Binance para operar en tiempo real. Los resultados mostraron un aumento de rentabilidad de 66.30% a 82.78%, pero con mayor riesgo, drawdown de 9.81% a 19.82%. La conclusión destaca el éxito del agente en adaptarse a condiciones de mercado y su potencial para trading en tiempo real, aunque requiere mejoras en la gestión de riesgos.

Palabras claves: aprendizaje por refuerzo, binance, trading

ABSTRACT

The objective of this project was to develop an algorithmic trading strategy using reinforcement learning, applied to the Ethereum futures market on the Binance exchange platform. The method included: collecting historical 5-minute candlestick data via the Binance API, optimizing technical indicators using grid search, training an agent with data from 2021-2023, evaluating it through backtesting in 2023-2024, and integrating it into the Binance API for real-time trading. The results showed an increase in profitability from 66.30% to 82.78%, but with increased risk, a drawdown from 9.81% to 19.82%. The conclusion highlights the agent's success in adapting to market conditions and its potential for real-time trading, although it requires improvements in risk management.

Keywords: reinforcement learning, Binance, trading

INTRODUCCIÓN

Este proyecto titulado “Desarrollo de Estrategia de trading algorítmico usando aprendizaje por refuerzo”, consiste en el diseño de un agente basado en aprendizaje por refuerzo para desarrollar una estrategia de trading adaptable al mercado de futuros de Ethereum en Binance. El propósito central es incrementar la rentabilidad mientras se controlan los riesgos asociados a las operaciones en este mercado. El estudio utiliza las capacidades del aprendizaje por refuerzo para que el agente se adapte de forma dinámica a las variaciones del entorno de mercado.

El proyecto se estructura en cinco fases desde recolección y preparación de datos, hasta conexión con la API de Binance para operar en tiempo real.

- Fase 1: Recolección, procesamiento y almacenamiento de datos históricos de velas de 5 minutos de Ethereum en archivos CSV, obtenidos a través de la API de Binance.
- Fase 2: Optimización de indicadores técnicos mediante ajuste de hiperparámetros con búsqueda de rejilla con el fin de encontrar la combinación con mayor rentabilidad y menor riesgo.
- Fase 3: Entrenamiento del agente de aprendizaje por refuerzo usando el algoritmo PPO en un ambiente de mercado simulado.
- Fase 4: Evaluación del desempeño del agente con datos históricos no usados en entrenamiento calculando métricas de rentabilidad.
- Fase 5: Integración del agente entrenado con la API de Binance para operaciones en tiempo real.

Este proyecto busca optimizar operaciones con criptomonedas en el mercado de futuros y establecer las bases para investigar aplicaciones de inteligencia artificial en mercados financieros, contribuyendo al desarrollo de soluciones financieras basadas en las tecnologías de la información.

CAPÍTULO 1. FUNDAMENTACIÓN

1.1 Antecedentes

En los últimos años, el desarrollo de algoritmos para la automatización y ejecución de estrategias de trading ha ido en aumento, debido al avance de las tecnologías como las APIs que permiten conectar plataformas de intercambio con estos sistemas[1]. Esto ha permitido que surjan grandes sistemas que gestionen órdenes de compra y venta sin intervención humana, mejorando la eficiencia de estas estrategias y elimina los errores provocados por decisiones subjetivas y emocionales [1].

La inteligencia artificial se utiliza cada vez más en diversos campos de estudio, surgiendo la necesidad de emplearla en escenarios donde las decisiones deben tomarse de manera rápida y objetiva [2]. Las técnicas de inteligencia artificial aprendizaje por refuerzo y aprendizaje profundo, han demostrado ser eficaces en los procesos de toma de decisiones basado en datos y contextos históricos; permitiendo la creación de sistemas más precisos que puedan adaptarse a situaciones complejas y cambiantes [2]. Esta tecnología se ha vuelto muy importante en sectores que necesitan realizar diferentes actividades con rapidez y precisión, como la medicina, la robótica y los mercados financieros [2].

El comercio de criptomonedas es uno de los sectores donde la inteligencia artificial ha cobrado especial interés y relevancia, los mercados de criptomonedas poseen constantes variaciones de precios que generan un ambiente de alta volatilidad que puede llevar a tomar decisiones impulsivas debido al estrés generado por la incertidumbre, provocando pérdidas significativas para los comerciantes [3]. La naturaleza impredecible de las criptomonedas aumentan el riesgo de tomar decisiones emocionales influenciadas principalmente en el miedo y la avaricia [3]. La inteligencia artificial aplica modelos matemáticos y algoritmos para permitir el aprendizaje, tiene el potencial para cubrir estos efectos al procesar grandes volúmenes de datos de diferentes naturalezas, brindando recomendaciones más objetivas y menos susceptibles a la subjetividad humana [3].

En el contexto específico de los mercados de futuros de criptomonedas, la necesidad de sistemas automatizados se ha vuelto más crítica debido a la operación continua

del mercado las 24 horas del día, los 7 días de la semana [4]. Los inversores enfrentan limitaciones físicas y cognitivas para monitorear y responder a las fluctuaciones del mercado de manera constante, lo que puede resultar en pérdidas significativas u oportunidades perdidas [4]. Además, la naturaleza altamente volátil de estos mercados, combinada con el alto apalancamiento disponible en las plataformas de trading de futuros, aumenta significativamente el riesgo de pérdidas sustanciales cuando se opera sin un sistema bien definido y automatizado [4].

A pesar de la capacidad de automatizar por completo estrategias de trading, persiste el problema de que los sistemas se basan en parámetros predefinidos y estáticos [5]. Con la volatilidad de los mercados es común que los algoritmos dejen de ser eficientes y se generen múltiples entradas falsas que afectan significativamente al rendimiento de la estrategia, debido a que, los algoritmos clásicos de trading automático no se adaptan a las condiciones fluctuantes del mercado y necesitan ser ajustados constantemente de forma manual [5].

En Madrid, se desarrolló un sistema de optimización de algoritmos de comercio analizando dos activos de acciones en la bolsa de valores Iberdrola y Solaria, se llegó a la conclusión que Iberdrola se comportaba mejor con los algoritmos clásicos basados en bandas de bollinger, por otra parte, Solaria se ajustaba mejor a los algoritmos basados en machine learning [5].

En Latinoamérica específicamente en el país de Colombia se desarrolló un Modelo de aprendizaje reforzado aplicado al trading de Bitcoin, en el que se obtuvo un 5% de rentabilidad, en este proyecto se utilizó una estrategia simple de medias móviles, sin embargo, abre la posibilidad a innovar en estrategias con diferentes temporalidades y combinaciones de indicadores que permitan mejorar el rendimiento [6].

En México, se presenta el método GA-DRL para la generación de estrategias de inversión con acciones de los mercados de valores [7]. En primer lugar, integra el algoritmo GenPo-Sharpe para la preselección de activos financieros cuya finalidad es proporcionar un portafolio de inversión con un número reducido de activos, de acuerdo con las preferencias del inversionista, y balanceado en cuanto a su rendimiento y riesgo esperados [7]. En segundo lugar, integra un agente de

aprendizaje por refuerzo profundo para el rebalanceo diario del portafolio de inversiones obtenido en la primera etapa [7]. La concatenación de estas dos técnicas, que constituyen al método presentado, se provee como una alternativa adicional para la gestión activa de portafolios de inversión con resultados prometedores en acciones de las bolsas de valores de Nueva York y Brasil [7].

Ante las necesidades previamente expuestas, este proyecto busca cubrirlas mediante el desarrollo de una estrategia de trading algorítmico utilizando aprendizaje por refuerzo. Los datos históricos de velas de precio e indicadores técnicos y la capacidad de aprendizaje del agente permitirán que la estrategia desarrollada se pueda adaptar a las variantes de los mercados de futuros.

1.2 Descripción del Proyecto

En el mercado de futuros de criptomonedas de Binance, existe un sinnúmero de scripts que automatizan estrategias, ejecutando operaciones y gestionando estas posiciones abiertas, sin necesidad de intervención humana, la mayoría de estos scripts usan filtros de acuerdo con el valor de indicadores o acción del precio, sin embargo, estos filtros funcionan con valores estáticos, lo que provoca que no se adapten a la volatilidad del mercado [5].

Bajo esta premisa, este proyecto tiene como fin generar un algoritmo que sea capaz de adaptarse a la volatilidad y diversas condiciones del mercado de futuros, usando la técnica de machine learning llamada aprendizaje por refuerzo, aplicado a la criptomoneda Ethereum.

Este proyecto se dividirá en cinco fases, las cuales fueron adaptadas del proyecto Modelo de aprendizaje reforzado aplicado al trading de bitcoin [6]:

Fase de recolección de datos.

Usando los datos públicos de la API de Binance, mediante la librería `python-binance`, `pandas` y `pandas_ta`, se desarrollará un script de python que recolecte los datos históricos de velas de precio de temporalidad de 5 minutos de la criptomoneda Ethereum de los últimos tres años, almacenando los datos en archivos csv distribuidos por cada año, generando cuatro archivos, uno adicional que contenga la totalidad de datos, usando estos, se calcularán los datos de indicadores técnicos:

- medias móviles exponenciales de 10, 20 y 50 períodos.
- una media móvil simple de 200 períodos.
- dos rsi de 8 y 14 períodos.
- un macd con longitud rápida de 12, longitud lenta de 26 y línea de señal de 9.

Fase de optimización de indicadores técnicos.

En esta fase, se evaluará el resultado de usar los indicadores técnicos generados durante la fase de recolección de datos, realizando una evaluación de rendimiento inicial. Posteriormente se buscará la mejor combinación de parámetros de indicadores técnicos para mejorar el porcentaje de acierto de las señales de trading. Para esta fase se usará la técnica de ajuste de hiperparámetros búsqueda de rejilla para generar combinaciones de parámetros para configurar los indicadores técnicos, a cada combinación generada se aplicará backtesting, para finalmente identificar cuál es la combinación que genera entradas más acertadas, obteniendo una mayor rentabilidad, ratio de ganancias, ratio de sharpe y menor drawdown en comparación con la evaluación inicial.

Fase de entrenamiento de agente.

Durante esta fase, se implementará un modelo de aprendizaje por refuerzo que pueda aprender una política, la cual será la estrategia de trading. Simulando un entorno de mercado y definiendo un espacio de tres acciones posibles para el agente, abrir operaciones de compra y venta, o esperar, basado en los datos históricos de velas de precios y los indicadores optimizados, el agente se entrenará mediante un enfoque de prueba y error. Dando como resultado que el agente maximice las recompensas en forma de ganancias por operaciones exitosas.

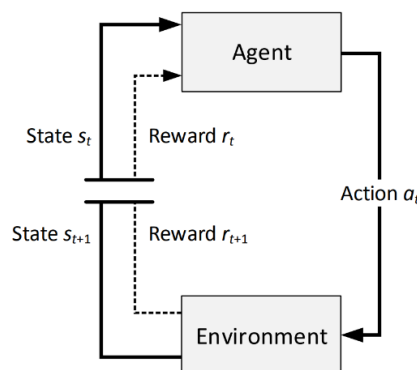


Ilustración 1 Aprendizaje por refuerzo Pröllochs & Feuerriegel, 2016.

Fase de evaluación de rendimiento.

En esta fase, se evaluará el desempeño del agente entrenado. Se aplicará backtesting usando datos históricos de un año que no fueron usados durante el entrenamiento para medir la rentabilidad generada, el promedio de ganancias y pérdidas, y además se obtendrán tres métricas de rentabilidad y riesgo:

1. Ratio de ganancias.
2. Máxima pérdida o max drawdown.
3. Ratio de sharpe.

Fase de integración al api de Binance.

Una vez que el agente ha sido entrenado y evaluado, se integrará a la API de Binance para operar en tiempo real en el mercado de futuros de la criptomoneda Ethereum. El agente decidirá cuando se enviarán órdenes de compra y venta, o cuando es mejor no operar para evitar riesgos; esta fase incluye la gestión de riesgo mediante órdenes de stop loss y take profit, y el registro del historial de las operaciones realizadas y sus resultados en un archivo csv.

1.3 Objetivos del Proyecto

Objetivo General

Desarrollar una estrategia de trading automatizada mediante el uso de la técnica de Machine Learning aprendizaje por refuerzo, aplicada al mercado de futuros de la criptomoneda Ethereum en la plataforma de intercambios Binance.

Objetivos Específicos

- Generar dataset de indicadores técnicos y velas de precios de criptomoneda Ethereum a través de la API de Binance.
- Optimizar parámetros de indicadores técnicos usando la técnica búsqueda de rejilla.
- Entrenar un agente de aprendizaje por refuerzo en un mercado simulado a partir del dataset generado.
- Evaluar el desempeño del agente aplicando backtesting comparando la ratio de operaciones exitosas.

1.4 Justificación del Proyecto

La implementación de una estrategia automatizada de trading basada en aprendizaje por refuerzo responde a la necesidad de superar las limitaciones de los sistemas tradicionales de trading algorítmico [8]. Los enfoques convencionales dependen de reglas estáticas y parámetros predefinidos, debido a esto resultan ineficaces frente a la naturaleza volátil del mercado de criptomonedas [8]. Investigaciones recientes han evidenciado que las estrategias que emplean aprendizaje por refuerzo logran superar a los métodos tradicionales en cuanto a rendimiento ajustado al riesgo, mostrando mejoras en la ratio de Sharpe [8].

Considerando un punto de vista técnico, el aprendizaje por refuerzo posee varias ventajas sobre otros métodos de machine learning aplicados al trading [9]. A diferencia de métodos supervisados que necesitan tener datos etiquetados previamente, el aprendizaje por refuerzo usa datos sin etiquetar para que un agente aprenda de manera autónoma mediante su interacción directa con un ambiente simulado, recibiendo una recompensa o penalización de acuerdo a los resultados de su acción realizada [9]. En el mercado de las criptomonedas, donde los patrones y niveles de precios cambian rápidamente, la capacidad de aprender directamente de un entorno es muy útil [9]. Diversos estudios previos a este proyecto han demostrado que los agentes entrenados con aprendizaje por refuerzo son capaces de desarrollar estrategias de trading adaptativas que son más rentables y menos riesgosas que los sistemas basados en filtros y reglas fijas [9].

Desde una perspectiva práctica, este proyecto surge por la necesidad de mejorar la gestión de riesgos en el trading de futuros de criptomonedas [9]. Los sistemas basados en aprendizaje por refuerzo pueden implementar estrategias de gestión de riesgos más avanzadas que los métodos anteriores, ya que son capaces de analizar y relacionar múltiples variables de forma simultánea y ajustar dinámicamente su política de trading según las condiciones cambiantes del mercado [9]. Hay pruebas concretas que demuestran que los sistemas de aprendizaje por refuerzo pueden reducir significativamente las pérdidas, especialmente las que comprometen un gran porcentaje del capital total, y lograr rendimientos constantes en comparación con estrategias que siguen reglas preestablecidas [9].

Por último, este proyecto se justifica por su capacidad para aportar al campo emergente del trading algorítmico en mercados de criptomonedas [10]. A pesar del creciente interés por las aplicaciones del aprendizaje por refuerzo en los mercados financieros, existen pocas investigaciones que aborden los desafíos específicos del mercado de futuros de criptomonedas utilizando indicadores técnicos [10]. El desarrollo de una estrategia exitosa no solo beneficiaría a los comerciantes individuales, sino que también enriquecería el conocimiento colectivo sobre la aplicación práctica de técnicas de inteligencia artificial en mercados de alta volatilidad como el de las criptomonedas [10].

1.5 Alcance del Proyecto

En este proyecto se desarrollará un algoritmo de trading automatizado usando aprendizaje por refuerzo para operar en el mercado de futuros de la criptomoneda Ethereum en la plataforma de intercambio Binance. Para simulación, entrenamiento, validación y pruebas se utilizarán datos históricos de velas de precio de 5 minutos de la criptomoneda Ethereum extraídos de la API de Binance, adicionalmente se añadirán datos de indicadores técnicos con parámetros optimizados mediante la técnica búsqueda de rejilla.

El rendimiento del agente de aprendizaje por refuerzo será evaluado mediante backtesting, analizando métricas de rentabilidad y riesgo: ratio de ganancias, máxima pérdida y ratio de Sharpe. Para la simulación del ambiente de mercado de futuros se considerará las comisiones de Binance al momento de abrir y cerrar una operación igual a 0.05% del tamaño de posición [11]. Finalmente, el algoritmo se integrará con la API de Binance para ejecutar operaciones en tiempo real, gestionando riesgos con órdenes de stop loss y take profit.

El proyecto no incluirá el desarrollo de interfaz gráfica. No se explorarán mercados distintos a los futuros de criptomonedas, ni se considerarán activos diferentes a Ethereum o temporalidades de velas de precio además de 5 minutos. El algoritmo no será entrenado operando en tiempo real, se limitará a ambientes simulados con datos históricos de velas de precio e indicadores técnicos. Los resultados obtenidos a lo largo de este proyecto no garantizan rentabilidad en mercados reales debido a la naturaleza volátil de las criptomonedas.

CAPÍTULO 2. MARCO TEÓRICO Y METODOLOGÍA DEL PROYECTO

2.1 Marco Conceptual

2.1.1 Aprendizaje por Refuerzo

El aprendizaje por refuerzo es una técnica de machine learning donde un agente aprende a tomar decisiones interactuando con un entorno, recibiendo recompensas o penalizaciones de acuerdo a las acciones realizadas, para maximizar una recompensa acumulativa [8]. Componentes principales [8]:

- Estado (S): Representación del entorno en un momento dado.
- Acción (A): Decisiones que puede tomar el agente.
- Recompensa (R): Señal numérica que indica el éxito de una acción.
- Política (π): Estrategia que sigue el agente para seleccionar acciones.

2.1.1.1 Aplicación al Trading

En el contexto del trading, los componentes se mapean de la siguiente manera [8]:

- Estado: Datos del mercado, indicadores técnicos, posición actual.
- Acciones: No operar, abrir compra, abrir venta.
- Recompensas: Rentabilidad de la operación o riesgo evitado.
- Política: Estrategia de trading aprendida.

2.1.2 Trading Algorítmico

El trading algorítmico, también conocido como algo trading, es un método de ejecución de órdenes de trading utilizando programas computarizados e instrucciones predefinidas que tienen en cuenta variables como tiempo, precio, volumen e indicadores técnicos [9]. Este enfoque automatizado permite ejecutar grandes bloques de órdenes de forma simultánea y sistemática sin intervención humana directa [9].

2.1.2.1 Backtesting

El backtesting o prueba retrospectiva es una técnica que permite evaluar el rendimiento de una estrategia de inversión mediante la simulación de operaciones usando datos históricos, replicando condiciones reales de mercado y características específicas como las comisiones [12]. Esta técnica permite validar estrategias

midiendo su viabilidad y rentabilidad antes de su implementación, permitiendo obtener métricas clave como retorno, porcentaje de aciertos, drawdown máximo y ratio de Sharpe [12].

2.1.3 Búsqueda de rejilla

La búsqueda de rejilla es una técnica de búsqueda exhaustiva, que evalúa combinaciones de hiperparámetros para identificar la combinación que ofrezca mejores resultados en un contexto establecido [13]. Explora un espacio predefinido, buscando la mejor solución dentro de rangos establecidos, normalmente, este proceso puede consumir muchos recursos computacionales [13]. En el aprendizaje por refuerzo, es crucial para ajustar parámetros como tasas de aprendizaje o arquitecturas de modelos para maximizar rendimiento y estabilidad [13].

2.1.4 Criptomoneda

Las criptomonedas son activos digitales basados en tecnología blockchain, funcionan como medio de intercambio descentralizado, utiliza criptografía para garantizar seguridad en transacciones y regular la emisión de nuevas unidades [14]. A diferencia de las monedas tradicionales, no tiene representación física y no son controlados por un Estado o entidad financiera central, facilitando transacciones directas entre usuarios [14]. Su valor varía según oferta, demanda y adopción en los mercados financieros [14].

2.1.4.1 Bitcoin (BTC)

Bitcoin fue desarrollado en 2009, su creador o grupo de creadores es conocido bajo el seudónimo Satoshi Nakamoto [15]. Opera como reserva de valor y método de pago, con un límite máximo de 21 millones de unidades, su blockchain emplea el método prueba de trabajo (PoW) para validar transacciones a través de minería [15]. Bitcoin suele considerarse como oro digital por su escasez y aceptación institucional; pese a su alta volatilidad, mantiene el liderazgo en capitalización de mercado y liquidez frente al resto de criptomonedas [15].

2.1.4.2 Ethereum (ETH)

Ethereum es la criptomoneda nativa de la red Ethereum, siendo la segunda más importante del mercado después de Bitcoin [16]. Es utilizada para pagar las

comisiones de transacción conocidas como gas fees, cuando se ejecutan contratos inteligentes o se realizan transferencias para realizar operaciones en la red Ethereum [16]. A diferencia de otras criptomonedas que solo sirven como medio de pago, ETH tiene un propósito más amplio al ser esencial para el funcionamiento de aplicaciones descentralizadas conocidas como dApps y participar en finanzas descentralizadas llamadas DeFi [16].

2.1.4.3 Tether (USDT)

Tether es una criptomoneda estable o stablecoin, creado por la empresa Tether en 2014 [17]. Posee todos los beneficios de la blockchain como, transparencia, velocidad y costes, pero a diferencia de otros cripto activos, su volatilidad es muy baja debido a que su valor está vinculado al dólar estadounidense, es decir 1 USDT tiene un valor igual o muy cercano a 1 USD [18].

2.1.5 Plataforma de intercambio de criptomonedas

También conocidas como brockers o exchanges, son infraestructuras centralizadas o descentralizadas que facilitan la compra y venta de activos digitales mediante mecanismos de emparejamiento de órdenes, determinando precios según la oferta y demanda [19]. Proporcionan liquidez y ofrecen diversos instrumentos financieros como spot, futuros y opciones [19].

2.1.5.1 Binance

Plataforma de intercambio de criptomonedas con mayor volumen de transacciones, permite operar en mercados spot, futuros y opciones en diferentes pares de Tether y USD Coin [20]. Binance ofrece un servicio de APIs REST y WebSocket, los cuales permiten acceso en tiempo real a datos de mercado y ejecución de órdenes con mínima latencia [21]. Además, proporciona datos históricos con múltiples temporalidades y un entorno de pruebas conocido como testnet para validar estrategias sin riesgo real [21].

2.1.6 Mercado de Futuros de Criptomonedas

Los futuros de criptomonedas son contratos financieros derivados donde las partes acuerdan comprar o vender un activo digital a un precio predeterminado en una

fecha futura, la cual puede estar definida o no, lo que se conoce como contrato perpetuo, la diferencia entre el precio acordado y el precio real de mercado definen el beneficio o pérdida obtenido [22]. Características principales [22]:

- Apalancamiento: Permite operar con una fracción del capital total.
- Margen de mantenimiento: Capital mínimo requerido para mantener posiciones abiertas.
- Liquidación: cierre total de la operación cuando las pérdidas superan el margen de mantenimiento.

2.1.6.1 Operación

En los mercados de futuros, una operación o posición se refiere a la ejecución de un contrato de compra o venta de activos con el objetivo de obtener beneficios [23]. El tipo de posición se basa en el movimiento esperado del mercado: una posición de compra o posición larga, implica adquirir un activo o contrato con la expectativa de que su valor aumente, la posición se cierra cuando se vende el total de la compra inicial; mientras que una posición de venta o posición corta, consiste en vender un activo no poseído, generalmente tomado en préstamo, con la anticipación de que su precio caiga [23].

2.1.6.2 Órdenes de Trading

Las órdenes de trading son instrucciones que detallan cómo y cuándo ejecutar transacciones en mercados financieros, especificando tipo de operación, cantidad y precio, permitiendo materializar decisiones de inversión en plataformas de intercambio [24]. Existen diferentes tipos de órdenes como mercado, stop loss y take profit, que permiten implementar estrategias y predefinir niveles para gestionar el riesgo, en el contexto de aprendizaje por refuerzo, las órdenes son acciones que el agente ejecuta según las observaciones del mercado [24].

2.1.6.2.1 Orden Stop Loss

La orden stop loss es una herramienta de gestión de riesgo que cierra una posición, parcial o completamente, cuando el precio alcanza un nivel predefinido; normalmente, se utiliza para limitar pérdidas o asegurar beneficios si el precio de un activo empieza a fluctuar en sentido opuesto a la tendencia [24]. Esta orden

ejecuta una transacción opuesta a la posición, es decir si es una posición de compra ejecuta una venta y viceversa [24].

2.1.6.2.2 Orden Take Profit

La orden take profit se ejecuta al alcanzar un precio objetivo en sentido favorable a la tendencia tomada en una operación, permite cerrar una posición parcial o completamente capturando beneficios, en ciertas ocasiones, cuando una operación ha tenido un mal rendimiento, se puede usar esta orden para obtener pérdidas mínimas [24].

2.1.6.3 Análisis Técnico

El análisis técnico en criptomonedas se basa en el estudio de patrones históricos y niveles de precio y volumen para predecir movimientos futuros [25]. Se fundamenta en el análisis velas japonesas, que contienen información de precio de apertura, cierre máximo y mínimo en un determinado intervalo de tiempo, junto con indicadores técnicos que permiten cuantificar la situación actual de mercado, permitiendo decidir a los inversores cuando entra o salir de una posición [25].

2.1.6.3.1 Velas japonesas

Son una representación gráfica de la variación de precio en una determinada temporalidad, muestran precio de apertura, cierre, máximo y mínimo en un intervalo temporal [26]. Originadas en el siglo XVIII, son estándar en mercados financieros por revelar patrones de comportamiento y sentimiento del mercado [26]. Su granularidad va desde segundos hasta meses, permite adaptarse a distintos enfoques de inversión, sea a corto, mediano o largo plazo, mientras que su formato estandarizado facilita el entrenamiento de modelos y la identificación de oportunidades de trading basados en la oferta y demanda. [26].

2.1.6.3.2 Indicador técnico

Los indicadores técnicos son herramientas utilizadas para orientar la toma de decisiones de inversión en diferentes condiciones de los mercados financieros [25]. Permiten identificar la fortaleza y calidad de una tendencia alcista o bajista, presentando la situación actual de mercado de forma cuantitativa y objetiva, se calculan a través de fórmulas matemáticas, normalmente tomando como fuente de

datos las velas de precio japonesas [25]. En este proyecto se utilizaron indicadores técnicos de diferentes tipos, lo cuáles son:

- Media Móvil Simple (SMA): Media aritmética de precios durante un número determinado de velas [25]:

$$SMA = \frac{Precio_1 + \dots + Precio_n}{n}$$

- Media Móvil Exponencial (EMA): Promedio de precios en un intervalo de tiempo que da más importancia a los datos recientes [25]:

$$Factor = \frac{2}{n + 1}$$

$$EMA = (Precio\ actual \times Factor) + (EMA\ anterior \times (1 - Factor))$$

- Índice de Fuerza Relativa (RSI): Es un indicador de tipo oscilador, ya que sus valores fluctúan entre 0 y 100, mide la velocidad y magnitud de los movimientos direccionales de precios [25]:

$$RS = \frac{Promedio\ de\ ganancias}{Promedio\ de\ pérdidas}$$

$$RSI = 100 - (100 / (1 + RS))$$

- Convergencias y Divergencias de medias móviles (MACD): Indicador de tendencia que muestra la relación entre dos medias móviles exponenciales [25]:

$$MACD = EMA(12) - EMA(26)$$

$$Signal = EMA(9)\ del\ MACD$$

2.1.7 Métricas de Evaluación

Las métricas de evaluación son indicadores cuantitativos que permiten medir de forma objetiva el rendimiento, eficacia y robustez de estrategias de trading, incluyen medidas como rentabilidad absoluta, rentabilidad ajustada por riesgo, drawdown máximo, ratio de operaciones ganadas y perdidas, volatilidad de retornos y frecuencia de operaciones, se debe evaluar en ambientes que presenten variedad de situaciones de mercados para brindar mayor fiabilidad [12]. Su cálculo riguroso es esencial para determinar la estabilidad y viabilidad de las estrategias, en el aprendizaje por refuerzo, complementan la recompensa acumulada del agente [12].

2.1.7.1 Ratio de Sharpe

Mide la rentabilidad ajustada al riesgo de una inversión, calculando el exceso de retorno sobre la tasa libre de riesgo, dividido por la desviación estándar de los rendimientos [12]. Se utiliza para evaluar la compensación de una inversión al riesgo asumido [12]:

$$\text{Ratio de Sharpe} = \frac{(Rp - Rf)}{\sigma p}$$

Donde:

- Rp: Retorno del portafolio.
- Rf: Tasa libre de riesgo.
- σp : Desviación estándar del portafolio.

2.1.7.2 Drawdown Máximo

La mayor pérdida de valor desde el punto más alto hasta el punto más bajo en el valor de una inversión o cartera, durante un período específico [12]. Se utiliza para medir el riesgo de pérdida potencial en una estrategia de inversión [12]:

$$MDD = \frac{(\text{Valor}_{\text{Máximo}} - \text{Valor}_{\text{Mínimo}})}{\text{Valor}_{\text{Máximo}}}$$

2.1.7.3 Ratio de ganancias

La ratio de ganancias es una métrica fundamental que cuantifica la relación entre el beneficio promedio de las operaciones exitosas y la pérdida promedio de las operaciones fallidas, evaluando el factor riesgo beneficio de una estrategia de trading [12]. Se calcula dividiendo la ganancia media por operación ganadora entre la pérdida media por operación perdedora, donde valores superiores a 1 indican un perfil favorable. Esta métrica requiere análisis conjunto con la tasa de aciertos para valorar su significado, ya que una ratio alta con baja frecuencia de éxitos puede ser engañoso [12].

2.1.8 Python.

Es un lenguaje de programación interpretado, de alto nivel y propósito general, se caracteriza por manejar una sintaxis clara y legible, además se destaca por su versatilidad, rápida escritura de código y facilidad para utilizarse programación

orientada a objetos [27]. En el contexto de este proyecto, Python se utiliza como lenguaje principal debido a sus librerías de machine learning, trading algorítmico y soporte para procesamiento de datos [27].

2.1.8.1 Gymnasium

Gymnasium es una librería de Python de código abierto que ofrece un entorno estandarizado para el desarrollo y evaluación de algoritmos de aprendizaje por refuerzo, permite la simulación de ambientes virtuales para que el agente interactúe mediante observaciones, acciones y recompensas, incluye herramientas para registrar episodios y visualizar resultados [28].

2.1.8.2 Pandas.

Biblioteca para manipulación y análisis de datos que proporciona estructuras de datos especializadas [29]:

- DataFrame: Estructura para datos tabulares, contiene filas y columnas.
- Series: Estructura para datos de una dimensión, compuesta por una columna o una fila.
- Pandas-ta: extensión de pandas especializada para el análisis técnico de mercados financieros, permite calcular más de 130 indicadores técnicos a partir de datos de velas de precios y parámetros específicos de cada indicador, además de permitir añadirles a las estructuras dataframe.

2.1.8.3 Python-binance.

Python-binance es una librería de Python para interactuar con la API de Binance, ofreciendo funcionalidades como la gestión de órdenes, obtención de datos históricos, y streaming de datos en tiempo real [30]. Sus funcionalidades permitirán desarrollar la estrategia de trading algorítmico en un ambiente real [30].

2.1.9 Archivos CSV

Los archivos CSV, valores separados por comas, son un formato de almacenamiento de datos ampliamente utilizado en análisis de datos [31]. Este formato permite almacenar grandes volúmenes de información en texto plano, con valores separados por comas, facilitando su manipulación y lectura en diversas herramientas de programación [31]. En el contexto del trading algorítmico, los

archivos CSV son útiles para importar y exportar datos históricos de precios, indicadores y resultados de operaciones [31].

2.2 Marco Teórico

2.2.1 Teoría de Ondas de Elliott

La Teoría de Ondas de Elliott, desarrollada por Ralph Nelson Elliott en los años 1930, postula que los mercados financieros presentan patrones cíclicos impulsados por la psicología colectiva de los inversores [32]. Estos patrones se caracterizan por secuencias de cinco ondas que avanzan en dirección de la tendencia general, seguidas de tres ondas correctivas que avanzan en la dirección opuesta [32]. Este modelo busca capturar el comportamiento recurrente del mercado, el cual es útil en el desarrollo de estrategias de trading algorítmico que pueden aprovechar estos ciclos predecibles para optimizar decisiones de compra y venta [32].

2.2.1.1 Estructura de las Ondas Principales

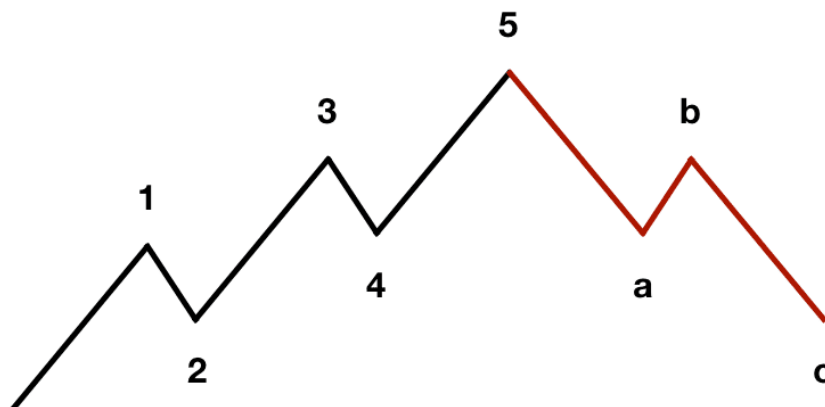


Ilustración 2 Ondas de elliot.

2.2.1.1.1 Ondas Impulsivas (1-5) [32]:

- Onda 1: Movimiento inicial, generalmente modesto.
- Onda 2: Retroceso que no supera el inicio de la onda 1.
- Onda 3: La más fuerte y extensa, nunca la más corta.
- Onda 4: Retroceso que no se superpone con la onda 1.
- Onda 5: Movimiento final del impulso, normalmente no supera la onda 3.

2.2.1.1.2 Ondas Correctivas (A-B-C) [32]:

- Onda A: Primera fase de la corrección.

- Onda B: Retroceso parcial contra la tendencia principal.
- Onda C: Movimiento final de la corrección.

2.2.2 Teoría de la fractalidad de los Mercados

La teoría de la fractalidad de los mercados financieros explica que estos presentan patrones similares en diferentes intervalos temporales, lo cual es clave para el desarrollo de modelos de aprendizaje por refuerzo en el trading [33]. Esta teoría sostiene que las estructuras y patrones de precios tienen características similares sin importar la escala temporal, lo que permite que un agente de aprendizaje por refuerzo aprenda políticas de patrones que se adapten a diferentes horizontes temporales [33]. Esta teoría se basa en los siguientes principios [34]:

- Autosimilitud: La naturaleza fractal de los mercados permite que patrones de comportamiento se repitan en diferentes escalas temporales, lo cual es útil para algoritmos que deben adaptarse a cambios en la temporalidad sin perder efectividad.
- Dependencia a largo plazo: Los niveles de precios tienen una memoria que explica que los eventos pasados pueden influir en el comportamiento futuro. Este principio es clave en el desarrollo de agentes de aprendizaje por refuerzo, los cuales pueden beneficiarse de esta dependencia temporal para anticipar movimientos de mercado.

2.3 Metodología del Proyecto

2.3.1 Metodología de Investigación

2.3.1.1 Contexto de la investigación

La presente investigación se sitúa en el contexto de los mercados financieros, específicamente en el mercado de futuros de la criptomoneda ethereum de la plataforma binance, la cual es un sitio de intercambio que nos permite operar con contratos de futuros en diversas criptomonedas [35]. Los mercados de futuros de criptomonedas, que operan las veinticuatro horas del día y los siete días de la semana, presentan desafíos significativos para los inversores, especialmente en la toma de decisiones [35]. La automatización de estrategias mediante inteligencia artificial ha surgido como una solución viable para abordar estos retos, reduciendo

la intervención humana y minimizando los sesgos emocionales que pueden afectar el rendimiento del comercio en mercados financieros [4].

En este contexto, la inteligencia artificial modelada a través del aprendizaje por refuerzo ha demostrado, por medio de diferentes estudios que anteceden esta investigación, ser eficiente para optimizar las operaciones de trading [7]. Los sistemas basados en esta tecnología permiten a los agentes responder de manera adaptativa a las fluctuaciones del mercado, ajustando sus estrategias en función de la retroalimentación continua del entorno financiero [7]. La investigación actual busca, por tanto, desarrollar un algoritmo automatizado de trading que pueda funcionar de manera eficiente en condiciones de alta volatilidad y con grandes volúmenes de datos en tiempo real.



Ilustración 3 Mercado de futuros de Ethereum en temporalidad 5 minutos

2.3.1.2 Enfoque de la investigación

El enfoque de esta investigación es cuantitativo, dado que se fundamenta en el análisis de datos históricos de precios y el rendimiento de indicadores técnicos [36]. El desarrollo del modelo, así como el entrenamiento del agente y su evaluación y validación requiere el cálculo de métricas de rentabilidad para comparar el desempeño de la estrategia y a otros métodos automatizados de trading [36].

Variable: en esta investigación se medirá el desempeño de la estrategia desarrollada a través del porcentaje de operaciones exitosas realizadas por el agente de aprendizaje por refuerzo, contrastando con el porcentaje de aciertos de las señales generadas durante el backtesting de la estrategia inicial.

2.3.1.3 Alcance de la investigación

El alcance de esta investigación es descriptivo, ya que se enfoca en observar y documentar el desempeño de un agente de aprendizaje por refuerzo en el desarrollo y ejecución de una estrategia de trading automatizada. A través de este enfoque, se busca evaluar la efectividad de la política aprendida por el agente en condiciones reales de mercado, considerando métricas de riesgo y rentabilidad como ratio de operaciones ganadas, ratio de Sharpe, caída máxima y rentabilidad absoluta [37].

2.3.1.4 Tipo de investigación

La investigación es de tipo experimental, ya que implica la creación de un agente de aprendizaje por refuerzo en un entorno de mercado de futuros simulado, manipulando variables como los parámetros de los indicadores técnicos para observar su efecto sobre el rendimiento de la estrategia de trading, permitiendo ajustar el modelo con base a los resultados obtenidos en los distintos escenarios [38].

2.3.1.5 Método de Investigación

Se aplica el método hipotético deductivo, partiendo de la hipótesis de que una estrategia basada en aprendizaje por refuerzo puede superar el desempeño de los sistemas tradicionales de trading algorítmico basadas en filtros estáticos, adaptándose a las condiciones cambiantes del mercado de criptomonedas; para finalmente evaluar su desempeño en un entorno simulado de mercado [39].

2.3.1.6 Población y muestra

La población de esta investigación está compuesta por los datos históricos de velas de precios, de temporalidad cinco minutos, del mercado de futuros de la criptomoneda Ethereum de la plataforma Binance de los últimos tres años, obteniendo 315360 registros. Para fines de esta investigación se usará la totalidad de la población como muestra, destinando dos años de datos para entrenamiento y un año para evaluación, lo que permitirá entrenar y evaluar el rendimiento del agente en un contexto representativo y realista. Esta selección permite obtener un conjunto de datos representativo que captura condiciones de mercados con movimientos alcistas, bajistas y laterales, con duraciones variables.

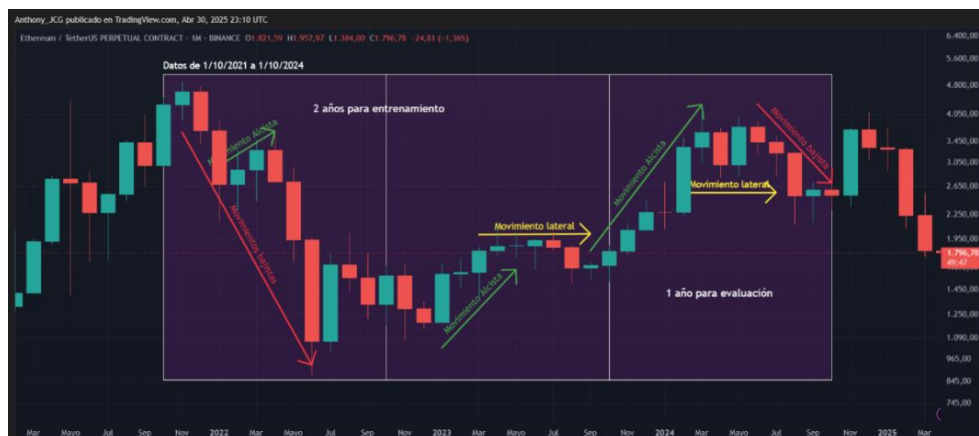


Ilustración 4 Gráfico mensual de los datos seleccionados como muestra

2.3.2 Técnicas e instrumentos de recolección de datos

La recolección de datos se realiza mediante la implementación de scripts automatizados utilizando la librería python-binance para acceder a los datos históricos. Los datos recolectados de velas de precio de temporalidad 5 minutos de Ethereum se procesan para añadir datos de indicadores técnicos calculados mediante la librería pandas_ta. Esta información se almacena en archivos csv empleando la librería pandas para su posterior análisis durante las fases de entrenamiento y evaluación del sistema.

2.3.3 Metodología de desarrollo

Para el desarrollo del algoritmo se sigue la metodología del proyecto Modelo de aprendizaje reforzado aplicado al trading de bitcoin, realizado en Medellín Colombia; adaptada a las necesidades específicas de esta investigación [6]. Las fases que este proyecto seguirá son [6]:

2.3.3.1 Fase de recolección de datos.

Se utilizarán datos históricos públicos de la plataforma Binance. Empleando la librería python-binance en conjunto con pandas, se diseñará un script en Python que almacenará en archivos csv los datos de velas de precio históricos en una temporalidad de 5 minutos de los últimos tres años. Una vez obtenido los datos, usando la librería pandas ta, se incluirán los indicadores técnicos: medias móviles exponenciales de 10, 20 y 50 períodos, una media móvil simple de 200 períodos, dos índices RSI de 8 y 14 períodos, y el MACD con una configuración de longitud rápida de 12, longitud lenta de 26 y una señal de 9.

2.3.3.2 Fase de optimización de indicadores técnicos.

Durante la optimización de indicadores técnicos, se analizarán los datos de indicadores técnicos obtenidos durante la recolección de datos. Para esto, se aplicará la técnica de búsqueda de rejilla para ajustar parámetros, probando diferentes configuraciones de los indicadores técnicos. Permitiendo identificar la combinación de parámetros que maximice la precisión de las señales de entrada de las operaciones generadas para maximizar las ganancias y reducir los riesgos .

2.3.3.3 Fase de entrenamiento de agente.

Se implementará un modelo de aprendizaje por refuerzo, diseñado para que un agente aprenda una política de trading rentable. Se simulará un ambiente virtual, usando la librería gymnasium, de mercado de futuros de la criptomoneda Ethereum usando los datos históricos de velas de precio y los indicadores técnicos optimizados, el agente podrá elegir una de tres opciones posibles, abrir operación de compra, abrir operación de venta o no operar; se entrenará en un proceso de prueba y error, obteniendo recompensas o penalizaciones según los resultados de las decisiones tomadas. El objetivo es aumentar la tasa de operaciones acertadas, maximizando la rentabilidad.

2.3.3.4 Fase de evaluación de rendimiento.

El objetivo de esta etapa es evaluar el desempeño del agente entrenado aplicando backtesting, utilizando datos de velas que no fueron usados durante la fase de entrenamiento. Para medir el rendimiento, se calcularán métricas de rendimiento y riesgo ratio de ganancias, pérdida máxima alcanzada, y ratio de Sharpe, para finalmente compararlas con la evaluación de la estrategia inicial.

2.3.3.5 Fase de integración a la API de Binance.

Una vez que el agente ha sido entrenado y evaluado, se integrará a la API de Binance para operar en tiempo real en el mercado de futuros de Ethereum. Se desarrollará un sistema que envíe órdenes de compra y venta basado en las decisiones del agente, y cada vez que se abra una nueva operación envíe órdenes de stop loss y take profit para gestionar el riesgo y tomar beneficios. El sistema registrará los resultados de las operaciones realizadas en un archivo CSV.

CAPÍTULO 3. PROPUESTA

3.1 Fase de recolección de datos

3.1.1 Datos de la API de Binance

En esta fase se recolectaron datos históricos de velas de 5 minutos de la criptomoneda Ethereum en el mercado de futuros a través de la API de Binance usando un script de Python, los datos recolectados se organizaron en cuatro archivos csv, cada uno se utilizará en fases específicas de este proyecto que se describen a continuación:

Nombre archivo dataset	Número de datos	Intervalo	Uso	Fase
velas_ETHUSDT_5m_año_1.csv	105120	01/10/2021 – 30/09/2022	Entrenamiento	3
velas_ETHUSDT_5m_año_2.csv	105120	01/10/2022 – 30/09/2023	Entrenamiento	3
velas_ETHUSDT_5m_año_3.csv	105120	01/10/2023 – 30/09/2024	Evaluaciones finales	4
velas_ETHUSDT_5m.csv	315360	01/10/2021 – 30/09/2024	Evaluaciones iniciales.	2

Tabla 1 Archivos generados al recolectar datos.

Los datos de los primeros dos años, 2021-2022 y 2022-2023, se destinan al entrenamiento del modelo en la fase 3 de este proyecto. El tercer año, 2023-2024, se reserva para la fase 4 de evaluación del agente. Por su parte, el archivo que contiene la totalidad de datos de tres años se emplea en evaluaciones iniciales de la fase 2 para analizar el desempeño de una estrategia tradicional basada en indicadores fijos, antes y después de optimizar los parámetros de indicadores.

Como se muestra en la Ilustración 4, estos tres años de datos incluyen condiciones de mercado alcista, bajista y neutral, lo que permite que el agente aprenda pueda ser entrenado y evaluado en diferentes movimientos de los niveles de precios.

Cada uno de los archivos generados, contienen la siguiente información, distribuida en las siguientes columnas, que corresponden a datos proporcionados por el API de Binance directamente:

Columna	Descripción
Open	Precio de apertura de la vela.
High	Precio más alto alcanzado por la vela.
Low	Precio más bajo alcanzado por la vela.
Close	Precio de cierre de la vela.
Volumen	Cantidad total de dólares operados durante la vela de 5 minutos.
Close time	Marca de tiempo de cierre de la vela.

Tabla 2 Descripción de datos recolectados desde el api de Binance.

3.1.2 Cálculo de indicadores técnicos

A través de la librería pandas-ta, se calcularon indicadores técnicos y fueron añadidos al conjunto de datos. Como se describe en el marco conceptual de este proyecto, los indicadores seleccionados pertenecen a diferentes tipos, lo que resulta en que se pueda tener una perspectiva amplia del mercado, los datos calculados fueron distribuidos en las siguientes columnas:

Columna	Descripción
RSI Fast	Índice de fuerza relativa de 8 periodos.
RSI Slow	Índice de fuerza relativa de 14 periodos.
EMA Fast	Media móvil exponencial de 20 periodos.
SMA Slow	Media móvil simple de 200 periodos.
Stop Short	Nivel para limitación de pérdidas para operaciones de venta basado en atr.

Stop Long	Nivel para limitación de pérdidas para operaciones de compra basado en atr.
MACD Line	Linea de convergencia de media móvil.
MACD Signal	Linea señal de convergencia de media móvil.
EMA Medium	Media móvil exponencial de 50 periodos.
EMA Fastest	Media móvil exponencial de 10 periodos.

Tabla 3 Descripción de datos de indicadores técnicos calculados.

Además de recolectar y agrupar los datos, se usó funciones de la librería pandas, para realizar una limpieza de datos, eliminando posibles datos nulos que puedan generar errores en las siguientes fases.

3.2 Fase de optimización de indicadores técnicos

3.2.1 Búsqueda de combinaciones de parámetros

Para la optimización de los parámetros, se utilizó la técnica búsqueda de rejilla, o grid search, un método de validación exhaustiva que evalúa todas las combinaciones posibles de un conjunto de parámetros predefinidos. Fueron considerados dos valores posibles para ocho parámetros de indicadores técnicos, lo que resultó en un total de 256 combinaciones. A cada combinación se aplicó backtesting en un periodo de tres años, utilizando los datos recolectados generado en la fase anterior.

Al finalizar las iteraciones, se obtuvieron los siguientes cambios en los parámetros:

Parámetro de Indicador	Antes	Después
RSI Fast	8	5
RSI Slow	14	14
EMA Fast	20	20
SMA Slow	200	200

MACD Fast	12	9
MACD Slow	26	20
MACD Signal	9	5
Tamaño mínimo de vela	0.0%	0.0%

Tabla 4 Cambios en los parámetros antes y después de la búsqueda de rejilla.

Se puede observar que cuatro parámetros correspondientes a dos indicadores fueron modificados tras la búsqueda de rejilla, RSI Fast, paso de una longitud de 8 a 5 periodos, y MACD que su longitud rápida, lenta y de señal pasaron de 12, 26 y 9 a 9, 20 y 5 respectivamente, mientras que los restantes se mantuvieron sin cambios.

3.2.2 Comparación de resultados obtenidos

Con los nuevos parámetros encontrados se actualizaron los datos de los indicadores técnicos para aplicar backtesting y comparar los resultados entre los indicadores iniciales y los optimizados en un periodo de 3 años correspondiente a la totalidad de datos. Para simplificar los cálculos se simuló un balance inicial de 100 dólares en cada evaluación, al final de estas, se calcularon métricas de rentabilidad, ratio de operaciones ganadas, porcentaje de caída máximo y ratio de Sharpe, además se obtuvieron los balances mínimos, máximos y finales, y el porcentaje de rentabilidad absoluta:

- La rentabilidad total en el periodo de 3 años aumentó de 30.51% a 84.04% evidenciando las mejoras obtenidas por el ajuste de parámetros.

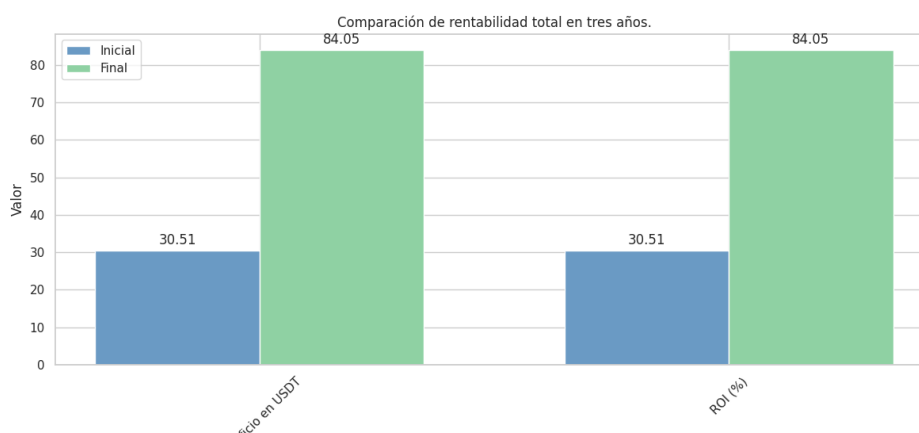


Ilustración 5 Gráfico comparativo de la rentabilidad obtenido en 3 años

- La ratio de operaciones ganadas pasó de 35.6% a 36.1% indicando que las señales fueron 0.5 veces más acertadas, reduciendo la ratio de operaciones perdidas de 64.4% a 63.9%:

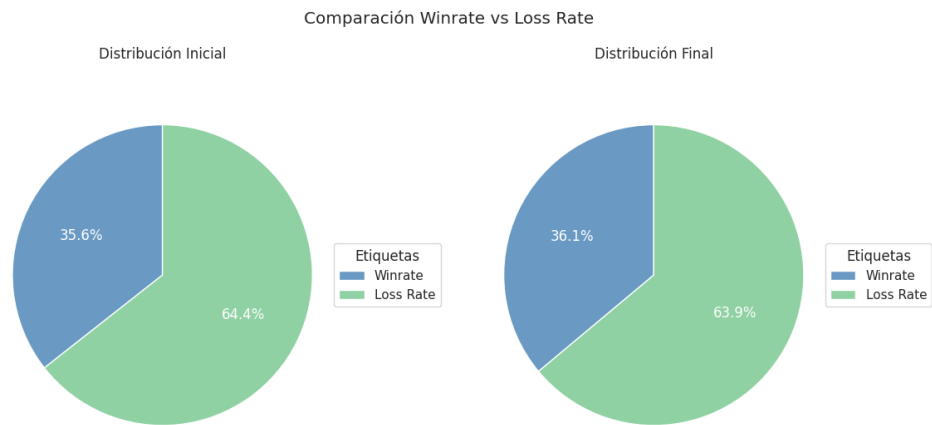


Ilustración 6 Distribución de la ratio de operaciones ganadas y perdidas

- Como se mencionó anteriormente, en cada backtesting se simuló un balance inicial de \$100, alcanzando en la primera evaluación un balance mínimo de \$83.09, un balance máximo de \$132.57, y un balance final de \$130.51, mientras que, con los parámetros optimizados los balances, mínimo, máximo y final fueron de \$90.26, \$195.09 y \$184.05, lo que indica que además de aumentar la rentabilidad, disminuyó el riesgo, ya que la disminución máxima fue de 10%.

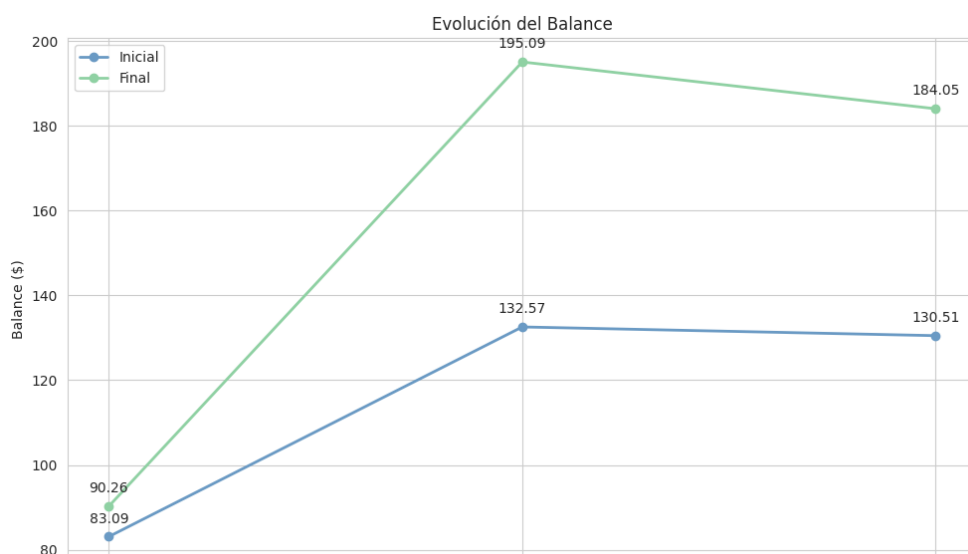


Ilustración 7 Balances máximos, mínimos y finales

- El drawdown máximo disminuyó de 36.26% a 35.19%, equivalente a que la pérdida potencial disminuyó en 1.07%. Por otra parte, la ratio de Sharpe aumentó un 40% pasando de 0.51 a 0.9, reflejando una mejor relación entre el riesgo asumido y el retorno obtenido.

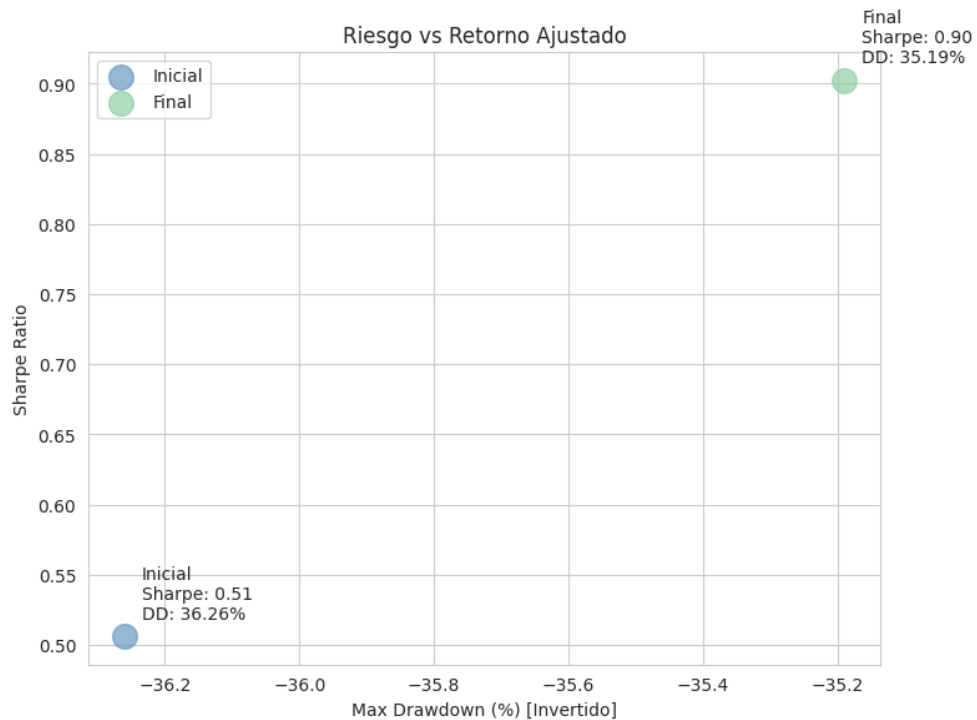


Ilustración 8 Variación de la ratio de Sharpe y drawdown máximo.

3.3 Fase de entrenamiento del agente

Para entrenar un agente que pueda desarrollar una estrategia de trading algorítmico en un entorno simulado se implementó un modelo de aprendizaje por refuerzo. El entorno utilizó los datos históricos recolectados y procesados en la primera fase de este proyecto, permitiendo al agente interactuar con las condiciones de mercado de dos años. El objetivo del entrenamiento fue que el agente aprenda una política que maximizara las recompensas, definidas en función de las ganancias obtenidas por entradas acertadas y pérdidas evitadas al decidir no operar, mediante un proceso iterativo de prueba y error.

Se empleó el algoritmo Proximal Policy Optimization PPO con una política basada en una red neuronal multicapa MlpPolicy para modelar las decisiones del agente, adaptada a un espacio de acciones discretas. El algoritmo PPO fue utilizado en su

versión enmascarable conocida como Maskable PPO, que permitió establecer reglas para impedir acciones inválidas como intentar abrir una nueva posición mientras otra estaba activa, permitiendo que el agente se enfoque en las decisiones de trading.

3.3.1 Ambiente virtual de simulación y espacio de observación

Para simular el ambiente virtual de mercado financiero se utilizaron los datos históricos previamente recolectados y procesados. El espacio de observación del agente incluyó una ventana retrospectiva de 288 pasos temporales, lo que equivale a 24 horas en velas de 5 minutos, cada uno con 15 características extraídas de los datos recolectados, presentes en las columnas: Open, High, Low, Close, Volumen, RSI Fast, RSI Slow, EMA Fast, SMA Slow, MACD Line, MACD Signal, EMA Medium, EMA Fastest, Stop Long y Stop Short, en la Tabla 2 y Tabla 3 se describe el contenido de cada característica.

Los valores fueron escalados y normalizados a un rango de -1 y +1 usando el método MinMaxScaler, que ajusta los valores originales transformándolos para que el valor menor sea igual a -1 y el mayor a +1 manteniendo la relación proporcional de los valores intermedios, para evitar distorsiones o problemas de interpretación por diferencias de magnitud [40].

Además, se añadieron nueve variables internas que complementan la observación para que el agente tenga una mejor perspectiva de la situación actual del mercado y del progreso de la inversión, estas variables fueron normalizadas bajo el mismo rango de -1 y +1, las cuales se describen en orden a continuación:

Variable	Descripción
sl_long_distance	Distancia porcentual entre el precio actual y el nivel de stop loss en posiciones de compra. Si el valor de la distancia es igual a 0% se normaliza a un valor de -1, mientras +1 representa una distancia máxima de 50%.

tp_long_distance	Distancia porcentual entre el precio actual y el nivel de take profit en posiciones de compra. Los valores siguen la misma escala y normalización que el anterior
sl_short_distance	Distancia porcentual entre el precio actual y el nivel de stop loss en posiciones de venta. Se sigue la misma normalización que los anteriores
tp_short_distance	Distancia porcentual entre el precio actual y el nivel de take profit en posiciones de venta. Normalizado bajo el mismo rango que los anteriores.
diff_balance	Diferencia porcentual del balance actual respecto al capital inicial. El valor se inicia en 0, cuando no hay ni pérdidas ni ganancias, +1 representa una rentabilidad igual o mayor a 100%, por otra parte, -1 indica la pérdida del 100% del capital.
win_rate_norm	Tasa de operaciones exitosas normalizada. 0 corresponde al 50% de aciertos, -1 al 0%, y +1 al 100% de aciertos.
drawdown	Reducción porcentual desde el balance más alto hasta el actual. -1 significa ausencia de drawdown, 0 el 50% de drawdown y +1 representa la pérdida total del balance.
vol_norm	Volatilidad histórica reciente normalizada. +1 indica máxima volatilidad y -1 ausencia de esta.
inact_norm	Tiempo de inactividad del agente, normalizado respecto a un límite máximo de 24 horas. +1 es igual a 24 horas de inactividad o más, -1 equivale a actividad constante.

Tabla 5 Descripción de variables internas de observación del agente

3.3.2 Espacio de acciones

El agente se desarrolló en un espacio de acciones discreto con tres opciones:

Opción	Acción
0	No operar.
1	Abrir posición de compra.
2	Abrir posición de venta.

Tabla 6 Espacio de acciones posibles del agente de aprendizaje por refuerzo

La implementación de Maskable PPO aseguró que el agente no seleccionara acciones inválidas, como abrir una posición adicional cuando ya había una activa. Esta máscara de acciones fue aplicada en el entrenamiento, evaluaciones, pruebas y ambientes reales, mantuvo la lógica del trading y optimizó el aprendizaje al limitar el espacio de decisiones a opciones válidas en cada estado.

3.3.3 Sistema de recompensas

El sistema de recompensas se diseñó para promover operaciones rentables y penalizar riesgos innecesarios sin motivar la pasividad o actividad extrema. Cuando el agente elige abrir una posición de compra o venta, el entorno simula de forma interna, sin mostrar nuevas observaciones al agente, la operación hasta alcanzar el stop loss o take profit. Si la operación es exitosa alcanzando el take profit, se otorga una recompensa proporcional a la ganancia; si, por el contrario, la operación trae pérdidas alcanzando el stop loss, se aplica una penalización proporcional a la pérdida conseguida. Además, se incluye una bonificación por alcanzar un nuevo máximo en el balance, y una penalización por una nueva caída máxima, estas bonificaciones y penalizaciones llevan a que el agente aprenda a realizar operaciones acertadas buscando acumular el mayor balance posible.

Si el agente decide no operar, el entorno simula internamente las posibles operaciones de compra y venta en ese punto. Si ambas operaciones tienen resultados

negativos, se recompensa por haber evitado pérdidas; si al menos una posición resulta rentable, se penaliza por haber perdido la oportunidad. Este sistema promueve un equilibrio entre acción e inacción, adaptando el comportamiento del agente a las condiciones del mercado. Finalmente, si han pasado más de 12 horas sin operar, se enmascara la acción 0 para que el agente decida abrir una posición sea de compra o venta, evitando así que aprenda una política pasiva.

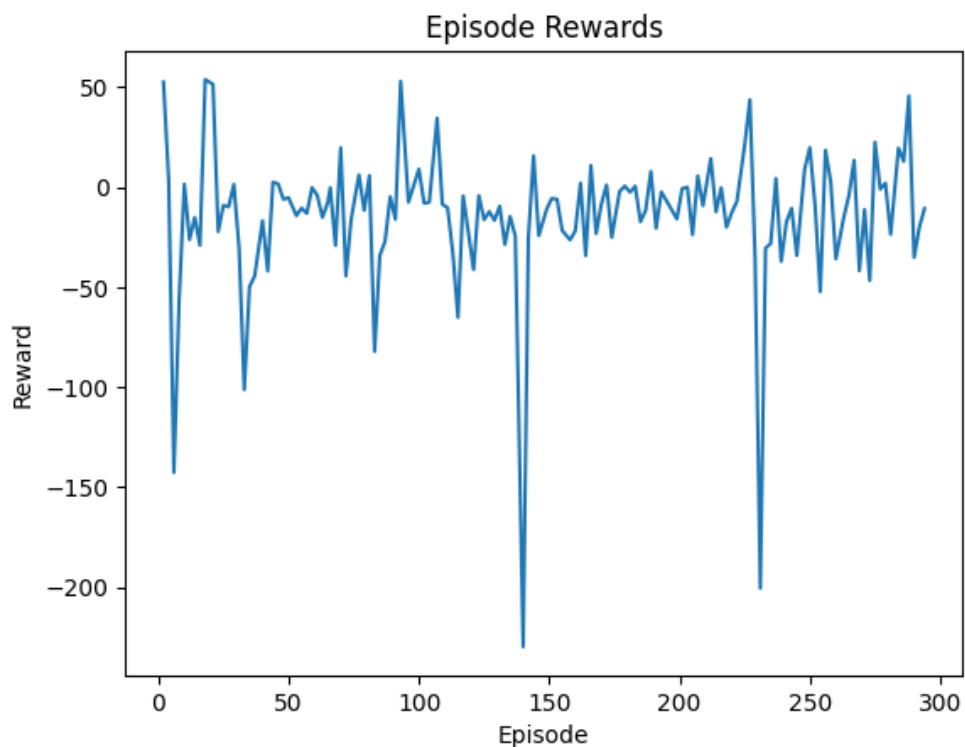


Ilustración 9 Recompensas obtenidas en cada episodio

3.3.4 Proceso de entrenamiento

El entrenamiento se llevó a cabo durante 137 episodios, cada uno con 4320 pasos temporales, lo que en velas de 5 minutos equivale a 15 días, al inicio de cada episodio se eligió una posición al azar del dataset para que así el agente pueda aprender en diversas condiciones del mercado, en total, el agente entrenó durante 591840 pasos, durando aproximadamente 4 horas. Se utilizó el algoritmo PPO con una tasa de aprendizaje de 5×10^{-6} y un factor de descuento gamma de 0.998 con el fin de aumentar la relevancia de recompensas futuras y generar una estrategia rentable a largo plazo. El objetivo fue maximizar las recompensas acumuladas,

vinculadas directamente a los resultados de las operaciones tomadas o evitadas, logrando aumentar la rentabilidad, disminuir el riesgo y obtener una mejor tasa de aciertos.

La ratio de operaciones exitosas promedio durante el entrenamiento fue de 35%, reflejando que, en la media, por cada 10 operaciones realizadas 3,5 fueron exitosas, alcanzando una tasa de operaciones ganadas mínima de 25% y una máxima de 45%.

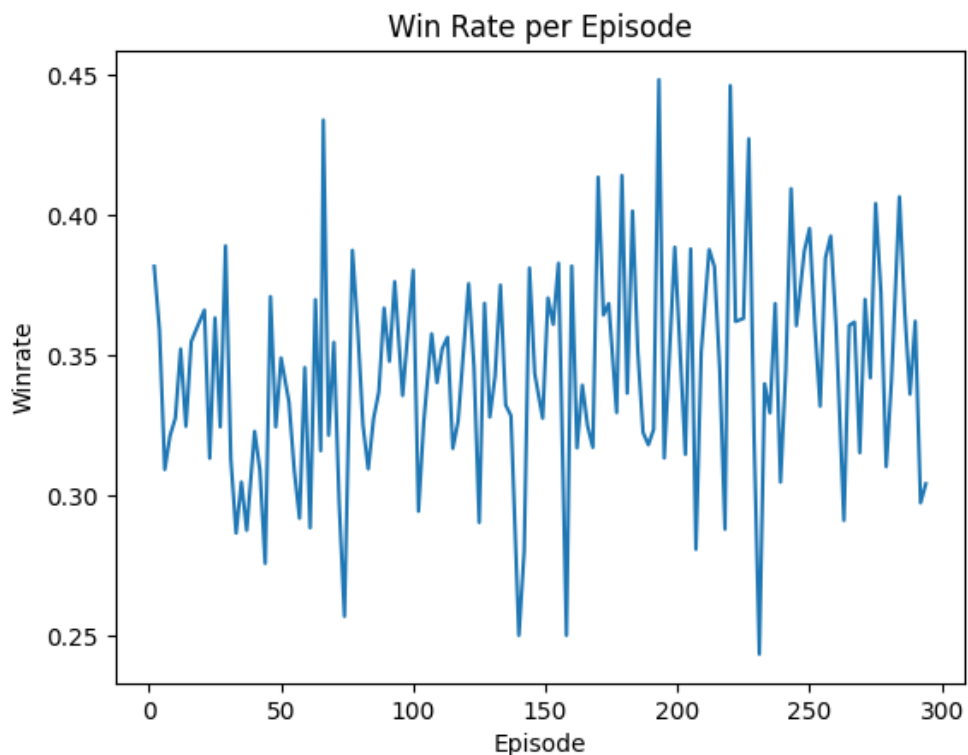


Ilustración 10 Ratio de operaciones ganadas en cada episodio

3.4 Fase de evaluación del agente

La fase de evaluación consiste en aplicar backtesting al agente entrenado utilizando datos históricos de un año no utilizados durante el entrenamiento. Este proceso mide de forma cuantitativa y objetiva a través de métricas de rentabilidad y riesgo la capacidad de predicción del modelo. Los datos utilizados para esta fase tienen la misma estructura que los datos utilizados durante el entrenamiento, al igual que en la fase dos, se simuló un balance inicial de 100 USDT.

En la siguiente tabla, se presentan los resultados obtenidos durante esta fase y las métricas calculadas a partir de esos resultados:

Métrica	Backtesting inicial	Backtesting agente
Balance final (USDT)	166.2951	182.7847
Balance mínimo (USDT)	94.2689	80.7245
Balance máximo (USDT)	177.3369	197.9666
Rentabilidad (%)	66.2951	82.7847
Ratio de Operaciones ganadas	37.7125	35.2062
Factor de Ganancias	2.0531	1.9336
Drawdown máximo	9.81	19.8231
Ratio de Sharpe	2.3530	1.3674
Operaciones totales	647	3079
Ganancia promedio	13.8940	15.8514
Pérdida promedio	6.7672	8.1980

Tabla 7 Resultados obtenidos de la estrategia inicial y agente RL

Se observa que el agente incrementó la rentabilidad absoluta, pasando de 66.30% a 82.78%, y el volumen de operaciones, de 647 a 3079 transacciones, lo que muestra que tiene mayor exposición al mercado. A pesar de que aumentó las ganancias, presentó un aumento del riesgo, el drawdown máximo se elevó de 9.81% a 19.82%, y la ratio de Sharpe disminuyó de 2.35 a 1.37, indicando una menor tasa de rentabilidad frente al riesgo asumido. La ganancia media aumentó de 13.89% a 15.85% lo que indica que en promedio las ganancias obtenidas por el agente son un 1.96% mayor, sin embargo, la pérdida media también creció de 6.77% a 8.20%, reflejando un aumento promedio en las pérdidas de 1.43%.

3.5 Fase de integración a la API de Binance

En esta fase, se integró el agente de aprendizaje por refuerzo entrenado a la API de Binance para operar en tiempo real en el mercado de futuros de la criptomoneda Ethereum y USDT. El objetivo fue habilitar al agente para tomar decisiones sobre el envío de órdenes para abrir posiciones de compra o venta, o la decisión de no operar para evitar riesgos. Además, cada nueva operación incluye gestión de riesgo mediante órdenes de stop loss y take profit, así como el registro del historial de operaciones y sus resultados en un archivo CSV.

3.5.1 Arquitectura del sistema de integración

El sistema de integración del agente de aprendizaje por refuerzo a la API de Binance se inicia con la recolección de datos de velas de 5 minutos del mercado de futuros de Ethereum. Usando los datos de velas obtenidos se calculan indicadores técnicos predefinidos desde la primera fase de este proyecto, para formar la observación del agente junto con variables internas del sistema, como saldo disponible, tiempo de inactividad y porcentajes de distancia entre el precio actual y los objetivos de pérdidas o ganancias. El agente, previamente entrenado y evaluado, analiza esta observación y selecciona una de tres acciones posibles: apertura de posición de compra, apertura de posición de venta, o abstenerse de operar para evitar riesgos.

Cuando el agente selecciona no operar para evitar riesgos, el sistema permanece en espera hasta el cierre de la vela actual de 5 minutos, una vez hay un nuevo cierre de vela se solicitan nuevos datos de mercado y empezar el proceso nuevamente. Si el agente decide abrir una posición, el sistema envía tres órdenes a través la API de Binance, una orden de apertura de la posición sea de compra o venta, una orden de stop loss para limitar pérdidas y take profit para asegurar ganancias.

Tras el cierre de cada operación, el sistema consulta la API de Binance para obtener detalles exactos del resultado, ganancia o pérdida, comisiones, tiempo de entrada y salida. Esta información se registra en un archivo CSV, añadiendo el campo de la dirección de la operación, es decir si fue de compra o venta. Luego de almacenar los datos, el sistema nuevamente esperará el cierre de la vela actual para el reinicio del ciclo.

Para evitar que el flujo del sistema se interrumpa debido a errores de solicitudes no exitosas, por fallos de red o respuestas no esperadas, se implementó un sistema de reintentos controlados, que almacena los errores en archivos de registros.

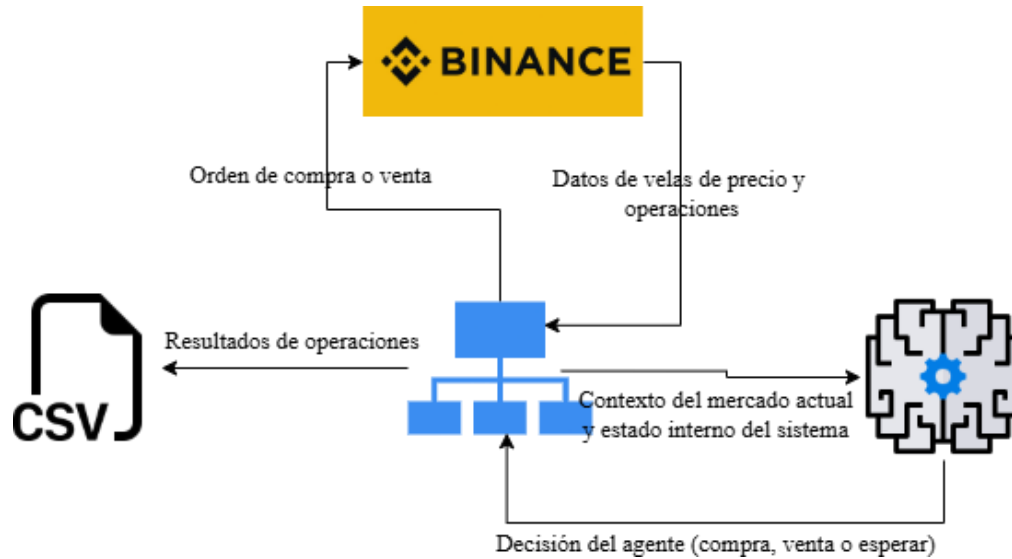


Ilustración 11 Esquema del sistema de integración entre el agente de RL y la API de Binance

CONCLUSIONES

- La generación del dataset a partir de la API de Binance permitió recopilar datos que incluyen velas de precios e indicadores técnicos de la criptomoneda Ethereum en el mercado de futuros. El dataset cubrió un período de 3 años, donde cada fila o registro corresponde a 5 minutos. Los datos fueron distribuidos en 3 archivos, donde cada archivo corresponde intervalos de un año, y un archivo general que almacena la totalidad de los datos.
- La búsqueda de rejilla optimizó 4 parámetros de indicadores técnicos, el índice de fuerza relativa (RSI) rápido, fue cambiado de 8 a 5, y los parámetros de convergencia y divergencia de las medias móviles (MACD) fueron cambiados totalmente, pasando de 12, 26 y 9 a 9, 20 y 5. Se evaluaron 256 combinaciones de parámetros, identificando la combinación de configuraciones óptimas, donde aumentó la rentabilidad en un periodo de 3 años, de un 30% a 86%. Disminuyó la caída máxima o drawdown de un 36,26% a 35,19%. Y mejoró la ratio de sharpe de 0.51 a 0.90, lo que indica que el beneficio obtenido cubre las pérdidas obtenidas.
- El agente de aprendizaje por refuerzo, entrenado en un entorno simulado basado en el dataset de Ethereum, se alcanzó una ratio de ganancias media de 35% durante 137 episodios, de los cuales cada uno corresponde 4320 pasos temporales, que en velas de 5 minutos corresponden a 15 días. El agente fue entrenando en total durante 591840 pasos temporales evaluando 137 condiciones de mercado diferentes, ya que en cada episodio se eligió un punto de inicio aleatorio.
- La evaluación del agente, usando el dataset del último año que no fue usado durante el entrenamiento, mostró que el agente logró una ratio de operaciones exitosas del 35%, junto con un factor de ganancias de 1.93 y una rentabilidad de 82,78% abriendo un total de 3079 operaciones, superando los resultados de la evaluación inicial, durante este mismo año obtuvo una rentabilidad de 66,29% junto con un ratio de operaciones exitosas del 37% y un factor de ganancias de 2.05 abriendo un total de 647 operaciones, lo que indica que el agente es menos conservador logrando maximizar la rentabilidad.

RECOMENDACIONES

- Utilizar rangos de datos más amplios, como 5 años y otras temporalidades como 15 minutos, o el uso de múltiples temporalidades combinadas, intentar combinar 1 semana y 1 hora o 1 día y 5 minutos, para obtener una mejor perspectiva del mercado. Además, considerar la inclusión de otro tipo de indicadores técnicos, como bandas de bollinger, para enriquecer el análisis de la situación actual de mercado. Buscar otra fuente de datos además de binance y otros activos como Bitcoin o Solana.
- Ejecutar la búsqueda de rejilla usando más combinaciones de parámetros, y considerando nuevas variables, como días u horarios en los que se opera o no, niveles de stop loss y take profit, límite de operaciones al día. Incorporar otras técnicas de ajustes de hiperparámetros como búsqueda aleatoria, considerar otras métricas como el factor de ganancias para elegir la mejor combinación.
- Ajustar la función de recompensa del agente para penalizar operaciones con alta volatilidad o drawdowns significativos mientras una operación esté abierta. Habilitar el entrenamiento del agente en tiempo real y con resultados de operaciones reales, tomando en cuenta las comisiones y otros factores que no se consideraron durante las simulaciones, como las comisiones por la tasa de financiación. Implementar técnicas de paralelización en el entrenamiento del modelo con TensorFlow o PyTorch, herramientas que permiten mejorar el procesamiento computacional y utilizar datasets más complejos.
- Implementar un sistema de monitoreo en tiempo real, para evaluar el desempeño del agente en el mercado de futuros, y poder visualizar su rendimiento en gráficas estadísticas, para así realizar las retroalimentaciones y ajustes respectivos.

REFERENCIAS

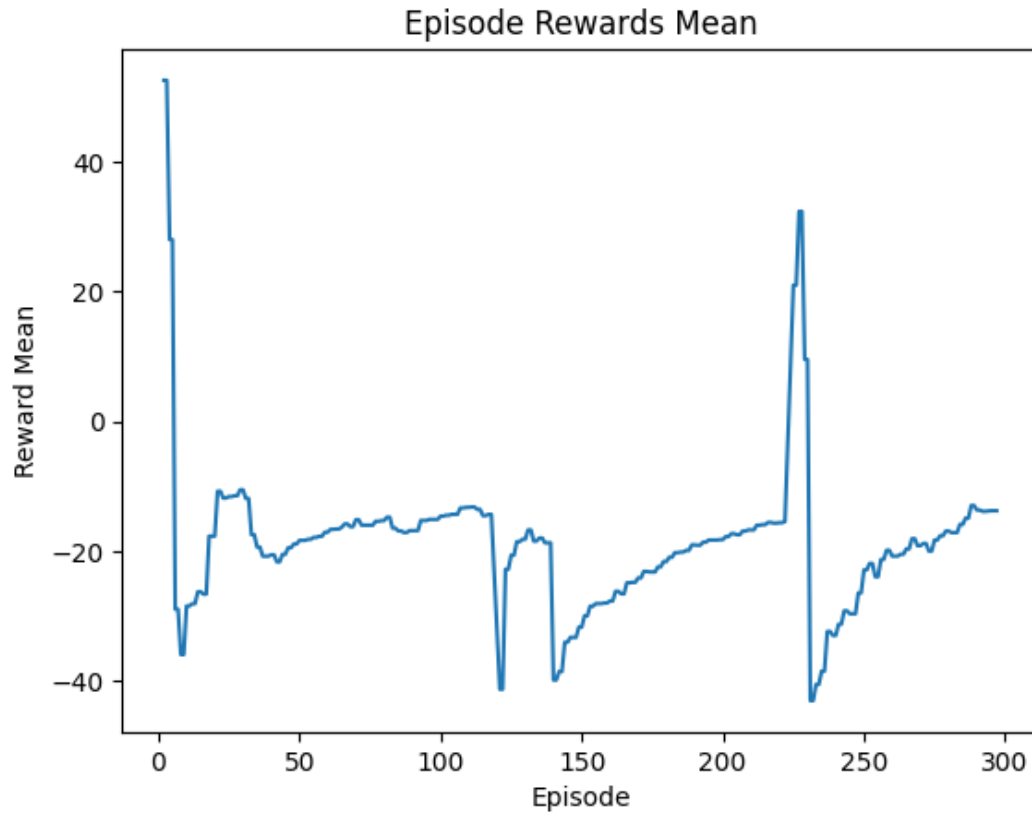
- [1] M. H. I. Gabriela, “El trading algorítmico en los mercados financieros”, PhD Thesis, 2018.
- [2] B. M. A, “Inteligencia Artificial - Margaret A. Boden”, en *Inteligencia Artificial*, 2016.
- [3] I. Paraskevopoulos, “Trading Algorítmico”, PhD Thesis, 2022.
- [4] J. C. Chávez y W. T. R. Ramos, “Trading Beneficios y Riesgos”, PhD Thesis, 2022.
- [5] M. P. Bravo, “Optimización de robots de trading”, PhD Thesis, 2023.
- [6] S. O. Morales, “Modelo de aprendizaje profundo reforzado aplicado al trading de Bitcoin”, PhD Thesis, Universidad EAFIT, 2022.
- [7] I. O. R. Hernandez, “Aprendizaje por Refuerzo Profundo para Trading”, PhD Thesis, Tecnológico Nacional de México, 2022.
- [8] S. A. G. Escobar, “Negociación algorítmica de acciones por medio de aprendizaje por refuerzo profundo”, PhD Thesis, 2021.
- [9] J. R. Siri y J. A. Serur, “Trading algorítmico: Una revolución en los mercados de capitales”, *Integr. Comer.*, núm. 44, pp. 204–215, 2018.
- [10] N. G. Nistal González Hugo, “Análisis y desarrollo de una estrategia de trading basada en diferencias de volatilidades”, Universidad Politécnica de Madrid, Madrid, 2020.
- [11] Binance, “Fees & Transactions Overview”. [En línea]. Disponible en: <https://www.binance.com/en-TR/fee/futureFee>
- [12] Z. Kakushadze y J. A. Serur, *151 Estrategias de Trading*. Ciudad de Buenos Aires, 2019. [En línea]. Disponible en: <https://arxiv.org/abs/1912.04492>
- [13] Petro Liashchynskyi y PavloLiashchynskyi, “Grid Search, Random Search, Genetic Algorithm: A Big Comparison for NAS”, 2019.
- [14] M. Cabrera Soto, “Criptomonedas: ¿qué son y qué pretenden ser?”, *Econ. Desarro.*, vol. 166, núm. 1, ene. 2022.
- [15] C. Domingo, *Todo lo que querías saber sobre BITCOIN, CRIPTOMONEDAS Y BLOCKCHAIN y no te atrevías a preguntar*. Barcelona: Editorial Planeta, 2018.
- [16] L. Girón, “Ethereum y las finanzas descentralizadas”, *Real. Empres.*, vol. 1, núm. 15, pp. 45–49, 2023, doi: 10.51378/reuca.v1i15.7722.
- [17] “Tether”. Consultado: el 17 de junio de 2025. [En línea]. Disponible en: <https://tether.to/es/>
- [18] I. S. Moreno, *Introducción al blockchain y criptomonedas en 100 preguntas*. Nowtilus, 2021.
- [19] Elisa Tasca, “CincoDías”, *Glosario Cripto*, Madrid, 2024.

- [20] Binance, “Binance About”. [En línea]. Disponible en: <https://www.binance.com/en/about>
- [21] Binance, “Binance Developer Center”. [En línea]. Disponible en: <https://developers.binance.com/>
- [22] A. Novales, “Midiendo el riesgo en mercados financieros”, PhD Thesis, 2017.
- [23] J. D. Tinoco y F. H. Trillo, *Futuros y opciones financieras: una introducción*. Editorial Limusa, 2002.
- [24] Rubén Villahermosa Chaves, *Trading e inversión para principiantes*, 2a ed. Madrid: RA-MA S.A. Editorial y Publicaciones, 2022.
- [25] F. R. Izquierdo, “Análisis técnico de los mercados financieros”, PhD Thesis, Universidad e Valladolid, 2021.
- [26] Orlando Mejía, “Todo sobre: Velas Japonesas”, 2006.
- [27] I. Challenger-Pérez, Y. Díaz-Ricardo, y R. Becerra-García, “El lenguaje de programación Python”, *Cienc. Holguín*, vol. 20, núm. 2, pp. 1–13, 2014.
- [28] Farama Foundation, “Gymnasium Documentation”. [En línea]. Disponible en: <https://gymnasium.farama.org/>
- [29] W. Mckinney, “Python-pandas: The library that marked a milestone in data analysis.”, 2011.
- [30] R. the Docs, “Documentación oficial de python-binance”. [En línea]. Disponible en: <https://python-binance.readthedocs.io/en/latest/index.html>
- [31] D. Golicher, “Introducción a bases de datos espaciales”, 2014.
- [32] J. L. M. Gordon, “Aspectos básicos de la teoría de la onda de Elliot”, en *Un enfoque múltiple de la economía española*, 2008, pp. 177–180.
- [33] J. M. S. Miguel, “La dimensión fractal en el mercado de capitales”, PhD Thesis, 2022.
- [34] L. G. V. Báez y J. V. A. Vera, “Antecedentes Fractales para Mercados Financieros”, *INCEPTUM*, vol. 5, pp. 263–292, 2010.
- [35] R. J. V. Montesdeoca, “Criptomonedas y el futuro del comercio internacional.”, PhD Thesis, 2024.
- [36] C. A. B. Novoa, *Investigación cuantitativa*. Fondo editorial Areandino, 2017. doi: 10.33132/9789585459007.
- [37] M. D. S. Tinta, “Tipos de Investigación Científica”, *Rev. Actual. Clínica*, vol. 9, pp. 621–624, 2011.
- [38] C. Ramos-Galarza, “Editorial: Diseños de investigación experimental”, *CienciAmérica*, vol. 10, núm. 1, pp. 1–7, 2021, doi: 10.33210/ca.v10i1.356.
- [39] A. M. Pujadas, “El método hipotético-deductivo de Karl Popper”, *Agenda Juárez Marginalidad Vulnerabilidad Suburbanización Cap. Andreu Marful*, vol. 14, pp. 16–20, 2024.

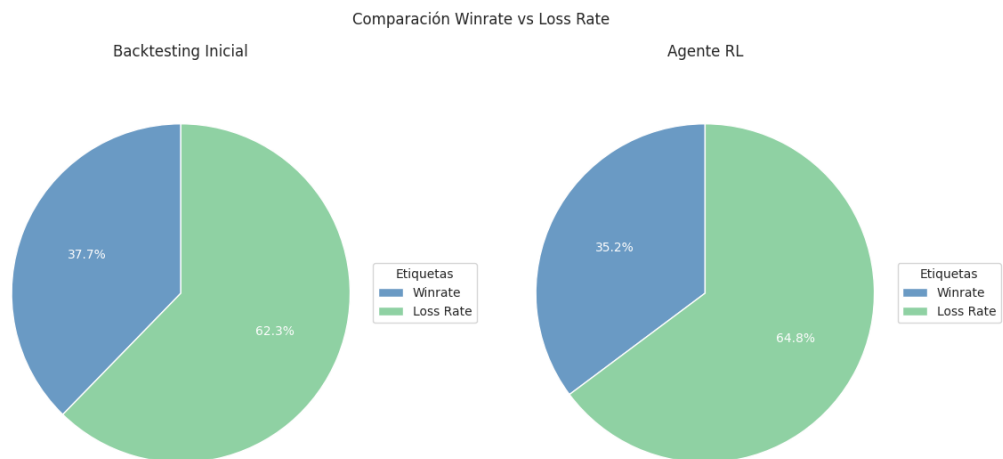
- [40] “MinMaxScaler — scikit-learn 1.7.0 documentation”. Consultado: el 19 de junio de 2025. [En línea]. Disponible en: <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.MinMaxScaler.html>

ANEXOS

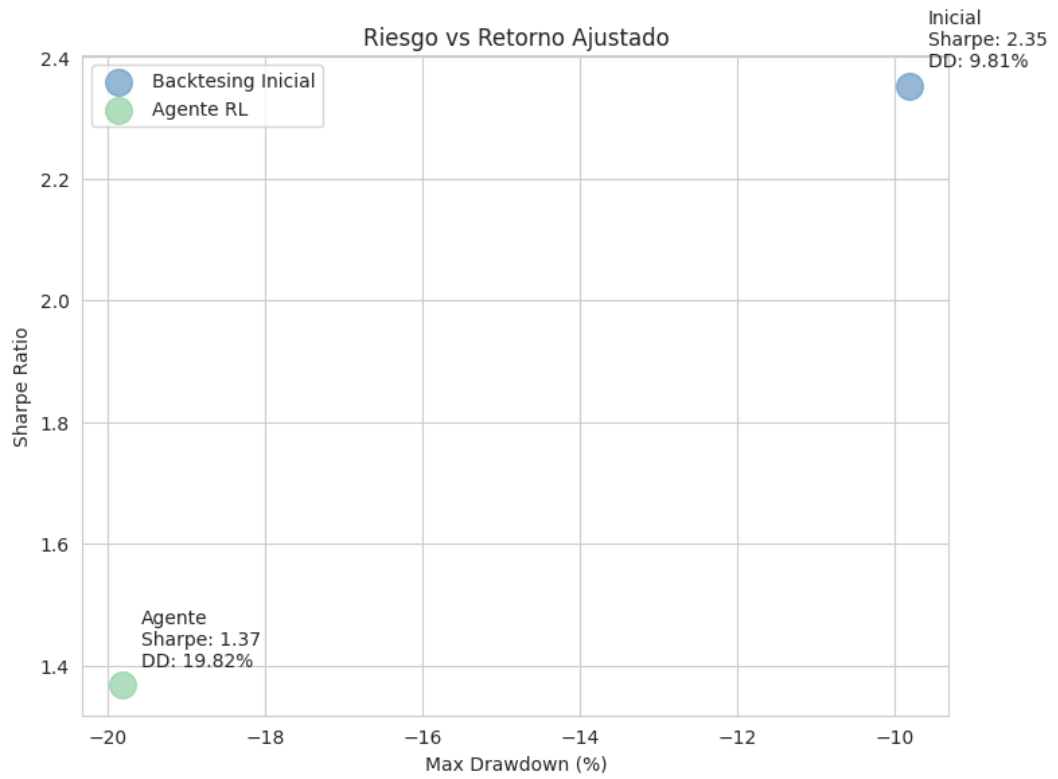
Anexo 1: Gráfico de la recompensa media obtenida durante la fase de entrenamiento del agente de aprendizaje por refuerzo.



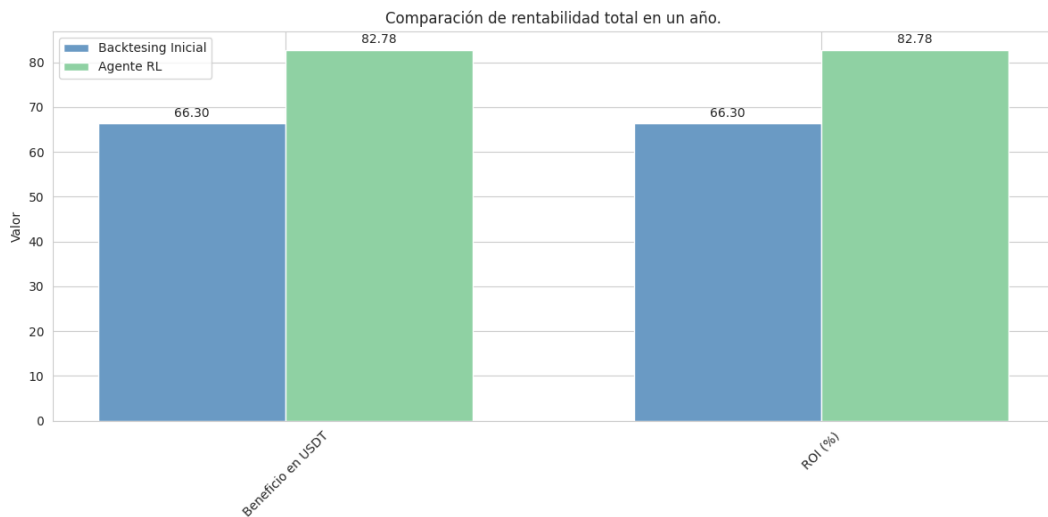
Anexo 2: Comparación de ratio de ganancias y pérdidas entre el backtesting aplicado a la estrategia inicial y el agente entrenado.



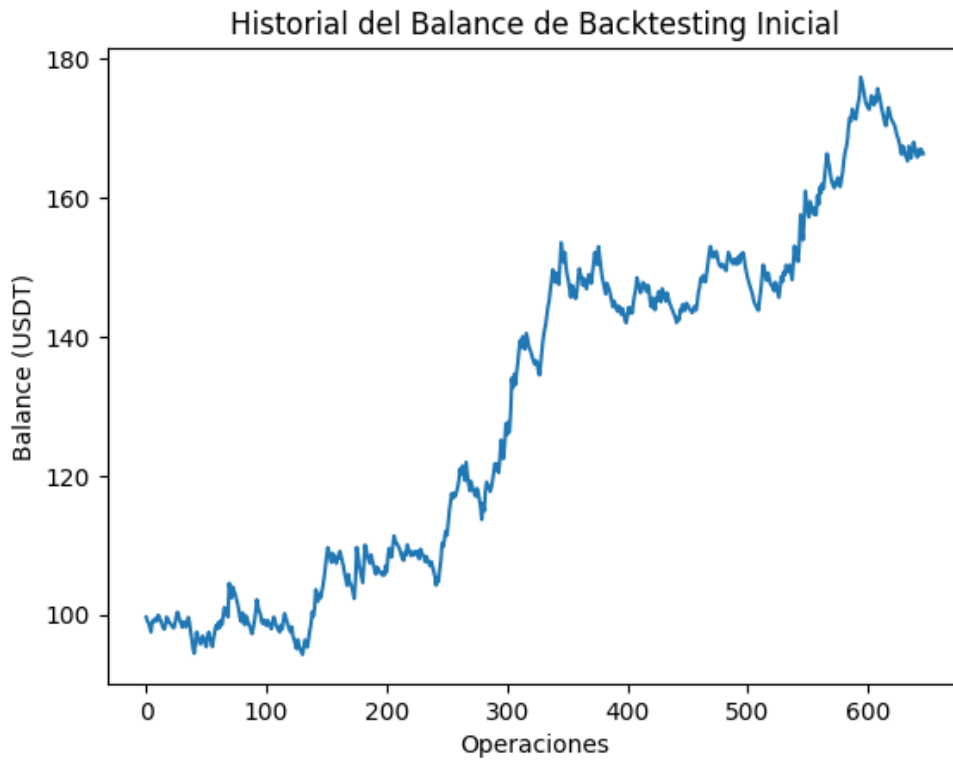
Anexo 3: Gráfico de dispersión de drawdown máximo y ratio de sharpe entre el backtesting inicial y el agente.



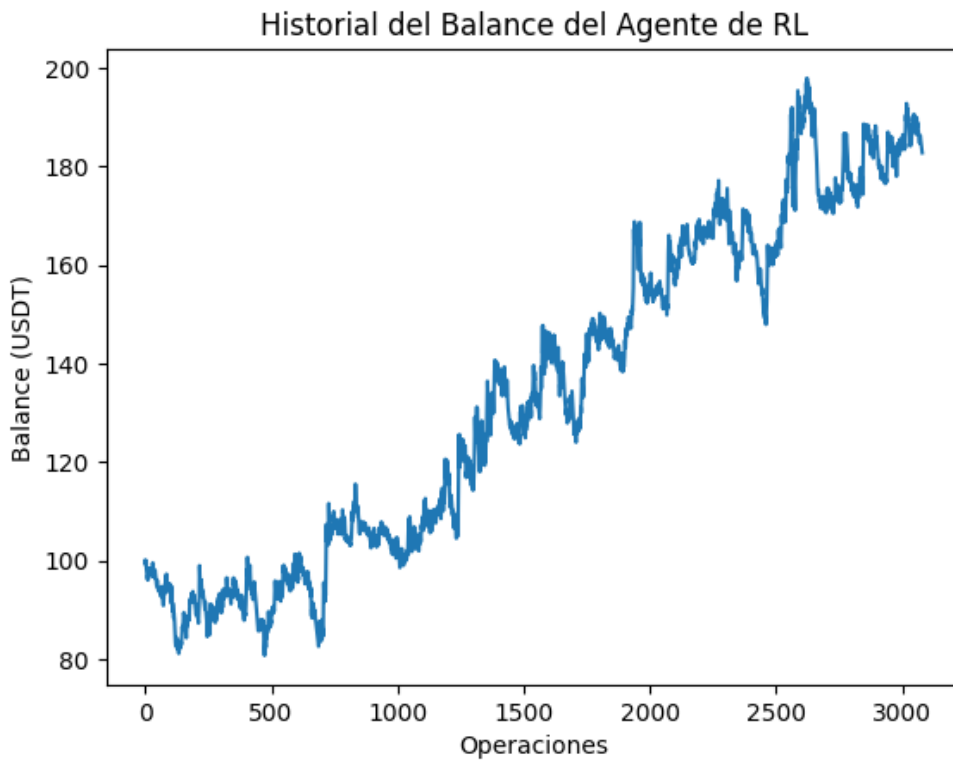
Anexo 4: Rentabilidad absoluta y beneficio de la evaluación inicial y final.



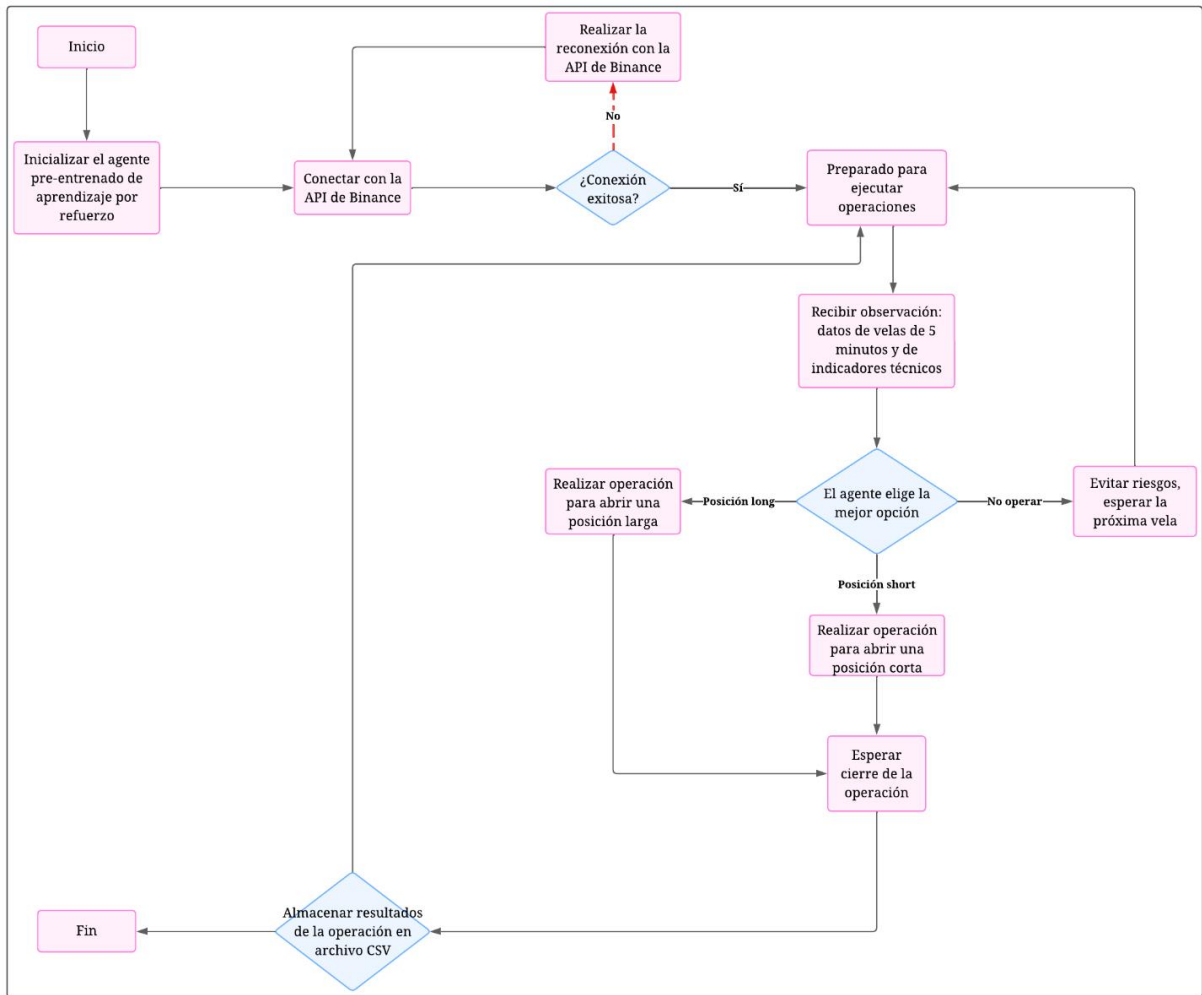
Anexo 5: Historial de balances durante el backtesting inicial durante 647 operaciones ejecutadas en un año.



Anexo 6: Historial de balance durante el backtesting del agente en 3079 operaciones ejecutadas en un año.



Anexo 7: Diagrama de flujo del sistema de integración a la API de Binance.



Anexo 8: Agente operando en tiempo real en la plataforma de binance

