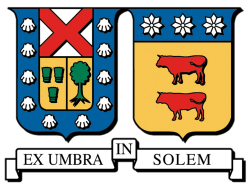


UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA
DEPARTAMENTO DE INDUSTRIAS

CONSENSO ADAPTATIVO A PARTIR DE MULTITUDES DE MODELOS: DESACUERDO INFORMACIONAL Y APRENDIZAJE POR REFUERZO EN LA CONSTRUCCIÓN DE PORTAFOLIOS

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL
Y AL GRADO DE MAGÍSTER EN CIENCIAS DE LA INGENIERÍA INDUSTRIAL

AUTOR
SEBASTIÁN ANDRÉS FLÁNDEZ HUERTA

PROFESOR GUÍA
WERNER KRISTJANPOLLER

CONSTANCIA DE VALIDACIÓN Y CONFIDENCIALIDAD DE MONOGRAFÍA A REPOSITORIO ACADÉMICO

1.- IDENTIFICACIÓN DEL TRABAJO ACADÉMICO

Tipo de monografía (marcar una opción): ☒ Memoria o trabajo de título ☒ Tesis de Postgrado

Título del trabajo: Consenso adaptativo a partir de multitudes de modelos: desacuerdo informacional y aprendizaje por refuerzo en la construcción de portafolios

Nombre del candidato(a): Sebastián Andrés Flández Huerta

Carrera / Grado: Ingeniería Civil Industrial / Magíster en Ciencias de la Ingeniería Industrial

Campus: Casa Central, Valparaíso **Departamento:** Departamento de Industrias

2.- VALIDACIÓN DEL PROFESOR GUÍA/DIRECTOR DE TESIS

Yo, Werner Kristjanpoller R., en mi calidad de profesor(a) guía/director(a) del trabajo académico mencionado anteriormente **DEJO CONSTANCIA** que:

- He revisado esta versión del documento y corresponde a la versión final aprobada del trabajo.
- El trabajo cumple con los requisitos académicos y de formato establecidos por la institución.

3.- EVALUACIÓN DE CONFIDENCIALIDAD POR PROPIEDAD INDUSTRIAL (marcar una opción)

☒ El trabajo **NO contiene** información que amerite confidencialidad y puede ser publicado de inmediato en repositorio con acceso abierto.

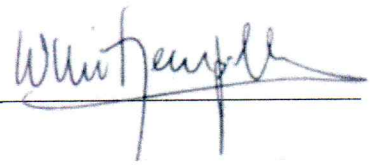
☐ El trabajo **CONTIENE** información con potenciales implicancias de propiedad industrial o intelectual y requiere un periodo de confidencialidad (**embargo**) por **(marcar una opción)**:

☐ 6 meses ☐ 12 meses ☐ 2 años ☐ 3 años ☐ 5 años ☐ 10 años

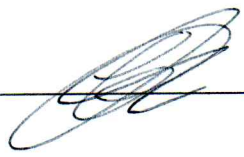
Fundamentación de la necesidad de confidencialidad (obligatorio si se solicita embargo):

4.- FIRMAS

Profesor(a) guía o director(a) de memoria o tesis:

Fecha: 10/07/2026 **Firma:** 

Estudiante o Candidato(a):

Fecha: 10/07/2026 **Firma:** 

Este formulario debe ser insertado como página 2 de la memoria o tesis, completado y firmado por estudiante y profesor(a) antes de la entrega en portal CRIS de Biblioteca USM.

Agradecimientos

A mi familia, por su apoyo incondicional a lo largo de todos estos años y por ser el sostén que hizo posible llegar hasta aquí. Gracias por la paciencia, por el aliento en los momentos difíciles y por creer en mí incluso cuando el camino se hacía largo.

A mis amigos, por acompañarme en este proceso, por las conversaciones que despejaron la mente y por el ánimo que volvió más liviana la carga.

A la Universidad Técnica Federico Santa María y, en particular, al Departamento de Industrias, por la oportunidad de formarme y por entregarme las herramientas y el espacio para desarrollar este trabajo.

A mi profesor guía, Werner Kristjanpoller, por la confianza, la orientación y la dedicación a lo largo de toda esta investigación.

Índice

Índice de figuras

Índice de tablas

Resumen Ejecutivo

La construcción de portafolios descansa sobre expectativas de retorno que en la práctica provienen de múltiples fuentes que discrepan entre sí; sin embargo, ese desacuerdo suele tratarse como ruido, la agregación de las señales como una regla estática y simétrica, y el riesgo como una propiedad de activos individuales. Este trabajo propone un marco que aborda esas tres limitaciones de forma conjunta. Un crowd artificial de nueve modelos predictivos heterogéneos genera pronósticos de retorno para cada activo y horizonte, a partir de los cuales se separan dos magnitudes: el consenso, que aporta el componente direccional, y el desacuerdo entre modelos, reinterpretado como riesgo informacional y expresado como una matriz de riesgo diagonal. Sobre esa base, un meta-agente basado en Double Deep Q-Learning aprende a reponderar el crowd según el régimen de mercado, sustituyendo la agregación fija por una dependiente del estado, y el problema de Markowitz se reformula en clave informacional, evitando la estimación de covarianzas de retornos. El marco se evalúa sobre cuarenta y seis acciones estadounidenses, tres horizontes de inversión y un esquema walk-forward con seis años de evaluación fuera de muestra que abarcan regímenes contrastantes, comparando además un pipeline de retornos crudos con uno que suaviza la variable objetivo mediante filtrado wavelet y winsorización. En el pipeline que preserva la señal predictiva, las estrategias informacionales superan a los benchmarks clásicos, incluida la exposición pasiva al mercado; y aunque la agregación estática resulta un rival exigente, al exhibir los mayores ratios de Sharpe agregados, la capa adaptativa iguala o supera su desempeño en el horizonte de ciento veinte días, donde la variante de convicción alcanza un ratio de Sharpe de 1,157 frente a 0,791 del índice de mercado; esa ventaja se acumula a lo largo del ciclo de evaluación y no en el episodio de estrés agudo, en el que la agregación estática resiste mejor. Los resultados son robustos a la elección de hiperparámetros del meta-agente y favorecen una ventana de entrenamiento deslizante frente a una expansiva, mientras que el suavizado de la variable objetivo deteriora el desempeño al erosionar la señal direccional. En conjunto, tratar el desacuerdo entre un crowd de modelos como riesgo informacional y aprender a reponderarlo de forma dependiente del estado constituye una vía viable para la construcción de portafolios informacionales, cuyo valor se manifiesta con

mayor nitidez en el horizonte largo, donde el contenido económico de la señal es más alto.

Palabras clave: combinación de pronósticos; desacuerdo entre modelos; riesgo informacional; aprendizaje por refuerzo profundo; optimización de portafolios; multitud artificial.

1 Introducción

La construcción de portafolios descansa, en última instancia, sobre expectativas acerca de los retornos futuros. En la práctica, esas expectativas no provienen de una única fuente, sino de la combinación de múltiples señales —analistas, modelos, indicadores— que con frecuencia discrepan entre sí. Esta discrepancia no constituye un mero artefacto de medición: la literatura de valoración de activos ha mostrado que la dispersión de creencias se encuentra incorporada en los precios y se relaciona con los retornos esperados posteriores (Diether et al., 2002; Yu, 2011), que dicha dispersión puede afectar la formación de precios cuando las opiniones pesimistas no logran incorporarse plenamente al mercado (Miller, 1977), y que el consenso de fuentes heterogéneas posee contenido económico al ordenar activos en portafolios (Barber et al., 2001). El desacuerdo, además, no se limita a encuestas o a precios agregados, sino que se manifiesta en decisiones concretas de asignación: ante un mismo evento público, inversionistas que interpretan el mundo mediante modelos subjetivos distintos rebalancean sus carteras en direcciones opuestas (Meeuwis et al., 2022). Conviven así dos magnitudes distintas en cualquier conjunto de pronósticos: una señal direccional agregada —el consenso— y la dispersión en torno a ella —el desacuerdo—. Sostenemos que tratar ambas como dimensiones separadas, y no reducir la segunda a ruido, es clave para construir portafolios informados.

La mayor parte de los enfoques existentes, sin embargo, trata estas dos magnitudes de manera asimétrica. El desacuerdo suele relegarse a una estadística descriptiva o de control, en lugar de constituirse en una dimensión de riesgo que alimente directamente la decisión de cartera. La agregación de las señales se realiza, además, mediante reglas estáticas y simétricas —típicamente un promedio simple—, bajo el supuesto implícito de que el modo de agregar es neutral respecto del resultado. Esta presunción no es inocua: la literatura de combinación de pronósticos ha documentado que las reglas simples resultan notablemente difíciles de superar mediante esquemas que estiman ponderaciones óptimas, de modo que la agregación estática opera como un punto de comparación exigente y no como un rival débil (Timmermann, 2006). Por último, el riesgo se concibe predominantemente como una propiedad de activos individuales y no como una propiedad relacional que emerge de cómo se combinan las fuentes. A ello

se añade una fragilidad bien documentada de la maquinaria de media-varianza, cuyos insumos estimados —en particular la matriz de covarianzas— se vuelven inestables justamente en los episodios de tensión en que más importan (Chow et al., 1999; Sikalo et al., 2022), dificultad que ha motivado la búsqueda de medidas de riesgo alternativas y de formulaciones más robustas frente a la incertidumbre de los retornos (Kamil et al., 2010).

Este trabajo propone un marco que aborda esas tres limitaciones de forma conjunta. Un crowd artificial de modelos predictivos heterogéneos genera señales de retorno sobre cada activo y horizonte, en línea con la evidencia de que combinar familias de modelos con distintas capacidades funcionales mejora la predicción de retornos fuera de muestra y tiende a dominar la selección de un único modelo ganador (Gu et al., 2020; Qian & Zhang, 2025; Wang & Chen, 2023), y con antecedentes que construyen señales de inversión a partir de la sabiduría de multitudes artificiales (Kristjanpoller et al., 2021). Un meta-agente basado en Double Deep Q-Learning aprende secuencialmente a reponderar a los miembros del crowd según el régimen de mercado (Mnih et al., 2015; van Hasselt et al., 2016; Zepeda et al., 2026), reemplazando la regla de agregación fija por una dependiente del estado, en consonancia con la necesidad de adaptación que imponen el ruido, la no estacionariedad y los cambios de régimen propios de las series financieras (Shahparast et al., 2023). Finalmente, las dos magnitudes del crowd se reinterpretan como los insumos de un problema tipo Markowitz: el consenso provee el vector de retornos esperados μ y el desacuerdo provee una matriz de riesgo diagonal Σ de naturaleza informacional, y no una covarianza estimada de retornos, en sintonía con formulaciones que buscan eludir la estimación directa de dichos insumos (Shalit, 2020). De este modo, el desacuerdo deja de ser una estadística secundaria para convertirse en una medida de riesgo que emerge de la relación entre las fuentes, y la regla de agregación deja de ser estática para volverse adaptativa y dependiente del estado.

Las contribuciones de este trabajo son las siguientes:

- Reconceptualizamos el desacuerdo entre un crowd de modelos heterogéneos como riesgo informacional: una magnitud que emerge de la relación entre las fuentes, y no de la historia de retornos del activo individual. Ello separa explícitamente el consenso (señal direccional, μ) de la dispersión entre fuentes (riesgo, Σ diagonal), en lugar de tratar la

dispersión como ruido descartable.

- Introducimos un meta-agente Double DQN que aprende a reponderar el crowd de forma dinámica según el régimen, reemplazando las reglas de agregación estáticas y simétricas por una agregación adaptativa.
- Reinterpretamos el problema de Markowitz en clave informacional: μ y Σ no son momentos estimados de los retornos, sino el consenso y el desacuerdo del crowd, lo que evita la estimación de covarianzas de retornos.
- Evaluamos empíricamente el marco sobre un universo de acciones estadounidenses en múltiples horizontes y bajo un esquema walk-forward, comparando agregación estática frente a adaptativa y contrastando contra benchmarks clásicos.

El resto de la memoria se organiza como sigue. La Sección 2 revisa la literatura sobre desacuerdo en la valoración de activos, agregación de pronósticos, aprendizaje por refuerzo en la construcción de portafolios y asignación relacional del riesgo. La Sección 3 describe los datos, el crowd predictivo, el meta-agente y la construcción de portafolios. La Sección 4 presenta los resultados y su discusión, la Sección 5 los análisis de sensibilidad, y la Sección 6 concluye.

2 Revisión de la literatura

2.1 Desacuerdo y heterogeneidad de creencias en la valoración de activos

La idea de que la divergencia de opiniones afecta el precio de equilibrio y el retorno esperado se remonta al modelo de creencias heterogéneas con restricciones a la venta en corto (Miller, 1977): cuando los inversionistas discrepan sobre el valor de un activo y los pesimistas no pueden vender corto de forma efectiva, el precio tiende a reflejar la valoración de la minoría más optimista más que la evaluación promedio del mercado, de modo que un mayor desacuerdo se asocia con precios más altos y, por ende, con menores retornos posteriores. La evidencia transversal es consistente con este mecanismo: las acciones con mayor dispersión en los pronósticos de utilidades de analistas obtienen retornos futuros significativamente más

bajos que acciones comparables, y dicha dispersión no puede interpretarse como un proxy de riesgo en sentido clásico, pues —pese a correlacionarse positivamente con beta, volatilidad de retornos e incertidumbre sobre ganancias— predice retornos futuros menores, y no mayores, lo que resulta incompatible con un marco de riesgo recompensado (Diether et al., 2002). Esta lógica se ha trasladado al nivel de portafolio, documentándose que la dispersión de creencias se relaciona negativamente con los retornos esperados posteriores, con un efecto más marcado en horizontes de uno a tres años, coherente con la lenta reversión a la media del propio desacuerdo (Yu, 2011). En paralelo, la teoría de valoración con creencias heterogéneas ha desarrollado formulaciones tratables en que los agentes discrepan sobre la dinámica de los fundamentales —crecimiento, volatilidad e intensidad de saltos—, mostrando que tales diferencias de creencias generan implicancias sistemáticas sobre los precios y los retornos de equilibrio (H. Chen et al., 2010).

El signo económico del desacuerdo no es, sin embargo, universal. Una tradición teórica anterior sostiene que la divergencia de creencias debería asociarse a una prima de riesgo positiva, de modo que un mayor desacuerdo elevaría los retornos esperados en lugar de reducirlos. La evidencia que distingue ambos regímenes apunta a las fricciones de mercado como factor determinante: el mecanismo de menores retornos opera allí donde las restricciones a la venta en corto impiden que las opiniones pesimistas se incorporen plenamente a los precios. El caso complementario ha sido examinado mediante una medida directa de desacuerdo entre dealers sobre las velocidades esperadas de prepago hipotecario —y no una proxy indirecta— en el mercado de mortgage-backed securities, altamente líquido y esencialmente libre de restricciones a la venta en corto (Carlin et al., 2014). En ese entorno, un mayor desacuerdo antecede a mayores retornos esperados, mayor volatilidad y mayor volumen de transacción, consistente con una prima positiva por desacuerdo, interpretada como riesgo de negociación o de selección adversa y distinta de los riesgos de mercado tradicionales, dado que el efecto persiste tras controlar por proxies usuales de riesgo financiero y macroeconómico. Un rasgo adicional de esta evidencia es que el desacuerdo no se reduce a volatilidad: esta última, por sí sola, no incrementa el volumen, y es solo en presencia de desacuerdo que la incertidumbre se traduce en mayor actividad de mercado (Carlin et al., 2014).

Para nuestros fines, lo relevante de esta línea es cuádruple: confirma que la dispersión de expectativas tiene contenido económico; que la creencia media y la dispersión en torno a ella constituyen información distinta; que el desacuerdo es separable de la volatilidad y no se reduce a ella; y que su signo económico no es una propiedad fija, sino que depende del mercado, de las fricciones presentes y de la estructura de riesgo del activo. Esta última constatación orienta nuestro tratamiento del desacuerdo. Nuestro marco no busca establecer el signo con que la dispersión entre fuentes se incorpora a los precios, sino interpretarla como riesgo informacional —una medida de fragilidad o incertidumbre sobre la calidad de la señal de consenso— a penalizar en la construcción del portafolio. Esta lectura es agnóstica respecto del signo de la prima: resulta compatible tanto con un régimen en que el desacuerdo se asocia a sobrevaloración y menores retornos (Diether et al., 2002; Miller, 1977; Yu, 2011) como con uno en que constituye una fuente de riesgo compensada con mayores retornos esperados (Carlin et al., 2014). En cualquiera de los dos casos, exigir mayor consenso direccional por unidad de desacuerdo constituye una decisión de asignación razonable.

La pertinencia del desacuerdo no se agota en su relación con precios y retornos agregados, sino que se manifiesta en decisiones concretas de cartera: ante un mismo evento público, inversionistas que interpretan la información mediante modelos subjetivos distintos ajustan la composición de sus portafolios en direcciones opuestas, aumentando o reduciendo su exposición accionaria según sus creencias (Meeuwis et al., 2022). Esta evidencia respalda que el desacuerdo es una magnitud relevante para la asignación, y no una mera estadística descriptiva. En esa misma dirección, trabajos de análisis de decisiones han propuesto incorporar el desacuerdo agregado como una restricción explícita sobre la composición del portafolio, en lugar de tratarlo como información secundaria (Fasth et al., 2016). La brecha que abordamos es, no obstante, doble: esta literatura mide el desacuerdo entre analistas humanos o intermediarios sobre fundamentales —contables o de valoración— y lo emplea como variable explicativa de retornos, volatilidad o volumen, o bien como restricción de un problema de decisión, sin trasladarlo a la construcción de cartera como insumo de riesgo; nosotros, en cambio, construimos el desacuerdo a partir de un crowd artificial de modelos predictivos heterogéneos —que, al igual que los dealers del mercado de mortgage-backed securities (Carlin et al., 2014), observan una

base común de información pero difieren en supuestos, arquitectura y sesgos inductivos— y lo reinterpretamos directamente como una dimensión de riesgo informacional que alimenta la optimización del portafolio.

2.2 Agregación de pronósticos y la sabiduría de las multitudes

La premisa de que combinar pronósticos heterogéneos mejora la calidad de la predicción agregada constituye un resultado central de la literatura de combinación de pronósticos (Timmermann, 2006). El argumento es de diversificación: salvo que pueda identificarse de antemano un modelo cuyos errores no puedan ser compensados por los de sus competidores, combinar las predicciones individuales domina a la selección de un único modelo, en estricta analogía con la diversificación de un portafolio financiero. Esta lógica es directamente aplicable a un crowd de modelos heterogéneos, cuyas diferencias en conjuntos de información, supuestos de modelación y forma funcional —lineal frente a no lineal, parámetros fijos frente a variables— son precisamente la fuente de las ganancias por combinación. Más relevante aún para el presente trabajo, esta misma literatura reconoce que, en numerosos problemas de decisión, el usuario del pronóstico no requiere únicamente una medida agregada de tendencia central, sino también información sobre la dispersión entre las predicciones individuales, entendida como una medida de la incertidumbre o el desacuerdo que rodea al consenso (Timmermann, 2006). Esta observación ofrece una justificación, desde la propia teoría de combinación, para tratar el consenso y el desacuerdo como dos magnitudes informacionales distintas, complementando la evidencia de valoración de activos discutida en la sección anterior.

En el contexto específico de inversión, el consenso de múltiples fuentes posee contenido económico. Comprar las acciones con consenso de recomendaciones más favorable y vender en corto las de consenso menos favorable genera retornos anormales brutos del orden de cuatro por ciento anual, que sin embargo no sobreviven de forma confiable a los costos de transacción dado el elevado turnover requerido (Barber et al., 2001). Este resultado encierra dos lecciones para nuestro marco: el consenso de fuentes heterogéneas es informativo, pero su valor depende críticamente de cómo se transforma en portafolio y de los costos que dicha implementación conlleva.

Un aspecto frecuentemente subestimado es la robustez empírica de las reglas de agregación más simples. La evidencia acumulada indica que esquemas elementales —el promedio simple, o ponderaciones inversamente proporcionales al error cuadrático medio que ignoran las correlaciones entre errores de pronóstico— resultan notablemente difíciles de superar mediante esquemas que estiman ponderaciones óptimas, fenómeno atribuido al error de estimación que estos últimos introducen cuando el número de modelos es grande respecto del tamaño muestral (Timmermann, 2006). Lejos de constituir un punto de comparación débil, la agregación estática y simétrica funciona por tanto como un benchmark exigente. Esta consideración delimita con precisión el aporte de una agregación adaptativa: el valor de aprender ponderaciones dependientes del estado no se da por supuesto, sino que debe demostrarse frente a reglas estáticas cuyo desempeño la literatura ha mostrado consistentemente sólido.

En paralelo, la literatura de aprendizaje automático en valoración de activos ha mostrado que los modelos no lineales —en particular árboles y redes neuronales— superan a las especificaciones lineales en la predicción fuera de muestra de retornos, ventaja atribuida a su capacidad para capturar interacciones entre predictores que las especificaciones lineales no incorporan (Gu et al., 2020). Resulta especialmente pertinente que, en ese mismo marco, una estrategia que combina por igual los portafolios construidos a partir de los distintos modelos alcance un desempeño comparable o superior al del mejor modelo individual, en línea con el principio clásico según el cual combinar tiende a dominar la selección de un único modelo ganador. Esta ventaja de los métodos no lineales se extiende a horizontes de inversión más largos (Qian & Zhang, 2025) y a la captura de relaciones de riesgo dinámicas que las especificaciones lineales tradicionales no recogen (Wang & Chen, 2023), lo que refuerza la conveniencia de combinar familias de modelos con distintas capacidades funcionales. En la misma dirección, el preprocesamiento de las series y los esquemas híbridos que integran etapas de reducción de ruido con modelos no lineales alteran de forma material el desempeño de los sistemas predictivos en finanzas (Chan Phooi M'ng & Mehralizadeh, 2016; Nobre & Neves, 2019). Más directamente relacionado con el presente trabajo, se han construido sistemas de trading que generan un crowd artificial de agentes y agregan sus señales mediante esquemas de sabiduría de multitudes artificiales, explotando la divergencia entre los agentes como fuente de información

(Kristjanpoller et al., 2021). Estas vertientes motivan el uso de un crowd artificial heterogéneo, pero dejan abierta la cuestión de cómo ponderar a sus miembros: la regla de agregación se asume, por lo general, fija, simétrica y exógena, y la dispersión entre las fuentes se descarta al promediar, en lugar de tratarse como una dimensión informacional. Cerrar esa brecha —aprender la ponderación y, a la vez, preservar la dispersión— motiva el meta-agente que se describe a continuación.

2.3 Aprendizaje por refuerzo en la construcción de portafolios

Los mercados financieros son ruidosos, no estacionarios y están sujetos a cambios de régimen, lo que motiva el uso de métodos capaces de adaptarse en el tiempo en lugar de reglas fijas; en esta línea se han propuesto clasificadores incrementales que aprenden en línea y manejan explícitamente el concept drift en la predicción bursátil (Shahparast et al., 2023). Entre los paradigmas adaptativos, el aprendizaje por refuerzo profundo ofrece un marco para tomar decisiones secuenciales dependientes del estado, con el Deep Q-Network como arquitectura fundacional (Mnih et al., 2015) —que estabiliza el entrenamiento mediante experience replay y una red objetivo separada— y su variante Double DQN como corrección del sesgo de sobreestimación de los valores de acción (van Hasselt et al., 2016). Este sesgo surge porque el operador máximo del Q-learning estándar utiliza la misma red para seleccionar y evaluar acciones, lo que tiende a favorecer estimaciones accidentalmente altas; Double DQN desacopla ambas operaciones y reduce así el sesgo, con mejoras que se vuelven especialmente relevantes en entornos ruidosos y no estacionarios como los mercados financieros.

En finanzas, estas técnicas se han aplicado a la construcción de portafolios, incluyendo extensiones recientes basadas en transformadores para el aprendizaje de valores de acción en mercados de criptomonedas (Zepeda et al., 2026). La mayoría de estos enfoques emplea el agente para fijar directamente los pesos de los activos. Nuestro uso es distinto y, en este sentido, complementario: el meta-agente no asigna pesos a los activos, sino que aprende a reponderar las fuentes de un ensamble predictivo, y la salida de ese ensamble —consenso y desacuerdo— se convierte luego en los insumos de un problema de optimización de cartera. El aprendizaje por refuerzo opera así sobre la regla de agregación, y no sobre la cartera de forma directa.

2.4 Asignación de riesgo, juegos cooperativos y riesgo relacional

Una tradición distinta, anclada en la teoría de juegos cooperativos, sostiene que el riesgo de un componente no es evaluable de forma aislada, sino por su contribución marginal dentro de las combinaciones posibles. Aplicada a carteras, esta perspectiva descompone el riesgo del portafolio en aportes por activo mediante el valor de Shapley (Shalit, 2021), y muestra además que la elección del principio de asignación —Shapley, nucleolus o cost gap— altera materialmente las conclusiones (Auer & Hiller, 2021). Asimismo, se ha mostrado que los portafolios de regresión permiten obtener portafolios tangentes mediante una formulación de mínimos cuadrados, evitando la estimación directa de los insumos clásicos del problema de media-varianza (Shalit, 2020), lo que conecta con nuestra posterior reinterpretación de dichos insumos. Trabajos más recientes introducen explícitamente la dimensión de ponderación: asignar pesos a los componentes, en lugar de agregarlos simétricamente, modifica la atribución del riesgo (Hiller, 2025b), al igual que imponer una estructura jerárquica sobre ellos (Hiller, 2025a). Estos estudios son de naturaleza ilustrativa o de simulación y no constituyen, por sí mismos, evidencia de mercado; su valor para nuestro trabajo es conceptual. Establecen que el riesgo es relacional —emerge de cómo se combinan los componentes— y que la forma de ponderarlos no es neutra. Nuestro enfoque comparte esta intuición, pero invierte su dirección: mientras esta literatura pondera activos de forma exógena, nuestro meta-agente aprende de forma endógena los pesos sobre los modelos del crowd, y es la dispersión entre esos modelos —no una covarianza de activos— la que define el riesgo.

3 Metodología

3.1 Datos y construcción de la variable objetivo

El análisis empírico se basa en un panel diario de precios y volumen correspondiente a un conjunto de cuarenta y seis acciones del mercado accionario estadounidense. Para cada activo se dispone de precios de apertura, máximo, mínimo, cierre, precio de cierre ajustado por dividendos y splits, volumen de negociación y capitalización bursátil. Estas variables permiten caracterizar tanto la dinámica de precios como la intensidad de transacción del mercado,

y constituyen los insumos a partir de los cuales se construyen los retornos financieros y las variables predictivas utilizadas en las etapas posteriores del estudio.

La base se organiza como un panel balanceado en el que cada activo posee observaciones completas a lo largo del horizonte temporal considerado, lo que permite que las comparaciones entre activos se realicen sobre un conjunto homogéneo de información. Antes de conformar el panel final, las observaciones con valores no positivos en precios, volumen o capitalización bursátil se tratan como no válidas. El panel definitivo se define entonces como la intersección de fechas y activos con información completa en precios, volumen, retornos logarítmicos, logaritmo del volumen y capitalización bursátil, de modo que el conjunto empírico resulta estrictamente común para todos los insumos requeridos por el sistema predictivo.

A partir de las series de precios ajustados se construyen los retornos logarítmicos diarios de cada activo. Para un activo i en el instante t, el retorno logarítmico se define como

$$r_{i,t} = \ln\left(P_{i,t}^{adj}\right) - \ln\left(P_{i,t-1}^{adj}\right), \tag{1}$$

donde $P_{i,t}^{adj}$ denota el precio ajustado del activo en el periodo t. Esta transformación se utiliza de forma habitual en el análisis de series financieras porque permite aproximar de manera aditiva los retornos acumulados en distintos horizontes y estabiliza el tratamiento estadístico de las series.

El horizonte temporal abarca más de dos décadas de observaciones diarias, lo que proporciona una base suficientemente amplia para representar distintos regímenes de mercado y episodios de volatilidad. Si bien la información disponible se extiende desde comienzos de la década de 2000 hasta comienzos de 2026 —margen que permite construir indicadores de tendencia de largo plazo sin pérdida de observaciones iniciales—, los ejercicios de entrenamiento y evaluación se concentran en subperiodos específicos, definidos de modo de garantizar comparabilidad entre modelos y coherencia en la implementación empírica.

La estrategia de estimación y evaluación no se organiza en un único bloque continuo, sino en una secuencia de ventanas móviles de tres años construidas para generar predicciones genuinamente fuera de muestra en distintos momentos del tiempo. Se consideran diez bloques consecutivos, correspondientes a los trienios que van de 2014–2016 a 2023–2025, desplazán-

dose un año en cada paso. En cada bloque, los dos primeros años se destinan al entrenamiento de los modelos predictivos y el tercer año se reserva exclusivamente para la evaluación fuera de muestra (out-of-sample, OOS), de manera que las ventanas de prueba corresponden sucesivamente a los años 2016 a 2025. Este esquema de tipo walk-forward permite analizar el comportamiento de los modelos en contextos macrofinancieros diferenciados, disponer de una serie extensa de ejercicios fuera de muestra consecutivos y reproducir un mecanismo realista de actualización en el que la información disponible se desplaza progresivamente en el tiempo. Al construir cada predicción únicamente con información observable hasta el momento de su emisión, el diseño preserva una disciplina temporal estricta y evita el uso de información futura. Esta disponibilidad de predicciones fuera de muestra en años consecutivos resulta además fundamental para la etapa posterior de aprendizaje del meta-agente secuencial, dado que permite construir una secuencia temporal de decisiones y recompensas sin incurrir en sesgos derivados del uso de información dentro de muestra.

Sobre esta estructura temporal se definen los horizontes de predicción, que determinan tanto la construcción de la variable objetivo como la frecuencia de decisión en las etapas posteriores. El estudio considera tres horizontes correspondientes a veinte, sesenta y ciento veinte días de negociación, equivalentes aproximadamente a un mes, un trimestre y medio año de mercado, lo que permite examinar las señales predictivas en escalas de corto, mediano y relativamente largo plazo. Para cada horizonte h, el retorno acumulado futuro del activo i se define como

$$R_{i,t}^{(h)} = \sum_{k=1}^{h} r_{i,t+k}, \tag{2}$$

es decir, el retorno logarítmico acumulado entre los periodos $t+1$ y $t+h$. Esta magnitud vincula la información disponible en el instante t con el comportamiento futuro del activo en el horizonte correspondiente y constituye la variable objetivo de los modelos predictivos desarrollados más adelante.

Finalmente, el análisis incorpora dos series de referencia destinadas a la etapa posterior de evaluación de portafolios: el índice S&P 500, utilizado como referencia de mercado, y una aproximación de la tasa libre de riesgo basada en bonos del Tesoro estadounidense a diez años.

Ambas series se encuentran disponibles para todo el horizonte considerado y se alinean con el calendario de negociación del panel accionario, de modo que las comparaciones de desempeño económico se realicen sobre fechas homogéneas.

3.2 Variables predictivas

La predicción de retornos mediante modelos estadísticos y de aprendizaje automático requiere un conjunto de variables capaz de resumir patrones dinámicos presentes en las series de precios y volumen. Con ese propósito, el conjunto inicial de variables explicativas combina un grupo reducido de variables base derivadas directamente del panel —el retorno logarítmico y el logaritmo del volumen— con un conjunto amplio de indicadores técnicos que capturan distintas dimensiones del comportamiento de mercado. En lugar de definir cada indicador de forma individual, la Tabla 1 organiza el conjunto completo según la dimensión que cada grupo busca representar, distinguiendo variables de tendencia, de momentum, de volatilidad y de volumen, junto con las variables ancla y una variable estructural de tamaño económico.

Tabla 1: Conjunto inicial de variables explicativas, agrupadas por dimensión de mercado.

Categoría	Variables incluidas	Dimensión de mercado capturada
Variables ancla	Retorno logarítmico del precio ajustado; logaritmo del volumen transado	Rentabilidad diaria y nivel de actividad; se preservan de forma obligatoria
Tendencia	Medias móviles simples y exponenciales de 5, 10, 20 y 60 días; banda central de Bollinger	Dirección y persistencia del precio en distintas escalas temporales
Momentum	Convergencia-divergencia de medias móviles y sus líneas de señal e histograma; momentum porcentual de 6, 12, 20 y 60 días; tasa de cambio de 10 y 20 días; índice de fuerza relativa de 14 días	Velocidad, fuerza y aceleración del movimiento de precios
Volatilidad	Ancho relativo de las bandas de Bollinger; rango verdadero promedio de 14 días; volatilidad realizada de 20 y 60 días	Dispersión y riesgo local de la serie de retornos
Volumen y presión	Acumulación-distribución de Williams; volumen en balance; variación diaria del logaritmo del volumen; volumen relativo de 20 días	Intensidad de negociación y presión compradora o vendedora
Tamaño económico	Capitalización bursátil	Escala estructural del activo; se incorpora de forma obligatoria

Las variables de tendencia resumen la dirección predominante del precio a distintas escalas, las de momentum cuantifican la velocidad y la fuerza relativa de sus movimientos, las de volatilidad aproximan la dispersión y el riesgo local de la serie, y las de volumen describen la intensidad de negociación y la presión entre compradores y vendedores. A este conjunto se añade la capitalización bursátil, que representa el valor total de mercado de cada empresa y

aproxima diferencias estructurales de tamaño que no se reflejan necesariamente en los indicadores derivados de precios y volumen. El retorno logarítmico y el logaritmo del volumen se tratan como variables ancla, esto es, características que se preservan obligatoriamente a lo largo del procedimiento de selección descrito a continuación.

Dado que el conjunto inicial contiene indicadores que capturan información parcialmente solapada, se aplica un procedimiento de selección orientado a reducir la redundancia entre variables. El criterio se basa en el coeficiente de correlación de Pearson y se calcula de forma individual por activo, utilizando exclusivamente la muestra de entrenamiento de cada subperiodo. Cuando dos variables presentan una correlación en valor absoluto mayor o igual a 0,80, se considera que aportan información redundante y se elimina una de ellas, conservando aquella con menor correlación promedio respecto del resto del conjunto. Las dos variables ancla quedan exentas de este filtro y se mantienen siempre, al igual que la capitalización bursátil, cuya inclusión se considera obligatoria por su carácter estructural.

Puesto que la selección anterior se realiza activo por activo, el conjunto final común se construye mediante un criterio de supervivencia transversal: se retienen únicamente las variables que superan el filtro de correlación en al menos el 80 % de los activos. De este modo se obtiene un vector de variables explicativas homogéneo para todo el universo, condición necesaria para el esquema de entrenamiento conjunto utilizado más adelante. Conviene precisar que este procedimiento de selección, así como el cálculo de la correlación, operan sobre las variables explicativas y no sobre la variable objetivo; el tratamiento aplicado a esta última, orientado a un propósito distinto, se describe en la subsección siguiente.

3.3 Tratamiento de las series y pipelines de entrenamiento

Además de la selección de variables explicativas, el estudio aplica un tratamiento específico sobre la variable objetivo con el fin de evaluar la sensibilidad de los modelos frente al ruido presente en los retornos. La motivación es que una fracción relevante de la variabilidad diaria de los retornos financieros corresponde a fluctuaciones transitorias de alta frecuencia que pueden dificultar la identificación de patrones predictivos, por lo que se construye una versión alternativa del objetivo en la que dicho ruido se atenúa antes del entrenamiento.

El primer componente de este tratamiento es una transformación basada en wavelets, herramienta que permite representar una serie temporal mediante componentes localizadas simultáneamente en tiempo y frecuencia y que ha sido empleada para analizar series no estacionarias en distintas escalas temporales (Pandelara et al., 2022). Desde una perspectiva estadística, la descomposición en el dominio wavelet ofrece un marco formal para la reducción de ruido, en la medida en que separa las componentes estructurales de baja frecuencia de las fluctuaciones de alta frecuencia atribuibles a ruido (Donoho, 1995), principio que ha sido aprovechado en sistemas de predicción financiera que combinan filtrado wavelet con modelos no lineales (Chan Phooi M'ng & Mehralizadeh, 2016). En la implementación de este estudio, la serie de retornos se descompone una sola vez, utilizando una wavelet de Daubechies de orden cuatro, en una componente de aproximación de baja frecuencia y una componente de detalle de alta frecuencia; la serie suavizada se reconstruye conservando únicamente la componente de aproximación. Esta especificación parsimoniosa elimina la fluctuación de mayor frecuencia sin imponer un suavizado excesivo, de modo que se preserva la estructura general de la dinámica de retornos.

El segundo componente es una winsorización aplicada sobre la serie ya filtrada, cuyo objetivo es limitar la influencia de observaciones extremas que podrían afectar de forma desproporcionada el entrenamiento. El procedimiento reemplaza los valores que se sitúan por debajo del percentil inferior y por encima del percentil superior por los valores correspondientes a dichos percentiles, fijados en el 1 % y el 99 % de la distribución. Mientras el filtrado wavelet atenúa el ruido de alta frecuencia, la winsorización acota los valores atípicos remanentes, complementando así la estabilización de la variable objetivo.

Estas dos transformaciones se aplican exclusivamente a la serie de retornos empleada para construir la variable objetivo, de forma separada para cada activo y restringida al tramo de entrenamiento de cada subperiodo, mientras que las variables explicativas se mantienen sin modificación para preservar la estructura informacional original de las señales de mercado. De manera deliberada, los retornos utilizados en la evaluación fuera de muestra permanecen en escala cruda, de modo que el desempeño predictivo y económico se mide siempre respecto de la dinámica efectivamente observada del mercado y se preserva la separación estricta entre entre-

namiento y evaluación. Los modelos de referencia basados en promedios móviles, que generan predicciones directamente sobre la ventana de evaluación, reciben en cambio un tratamiento causal anclado en cada fecha de decisión, descrito junto a su especificación en la subsección correspondiente.

Este diseño da lugar a dos versiones paralelas del objetivo de predicción y, en consecuencia, a dos pipelines de entrenamiento. En el primero, denominado RAW, los modelos se entrenan utilizando como objetivo los retornos acumulados en escala cruda. En el segundo, denominado WW, los modelos se entrenan sobre la versión filtrada y winsorizada de la serie. La existencia de ambos pipelines responde a la evidencia de que el preprocesamiento de las series altera de forma material el desempeño de los sistemas predictivos en finanzas (Nobre & Neves, 2019), y permite contrastar empíricamente si la reducción del ruido en la variable objetivo se traduce en mejoras de capacidad predictiva, manteniendo en todos los casos la evaluación fuera de muestra anclada a retornos observados sin procesamiento.

3.4 Crowd predictivo

El núcleo predictivo del marco propuesto es un crowd artificial de modelos heterogéneos que generan, de forma independiente, pronósticos de retornos acumulados futuros sobre cada activo y horizonte. La heterogeneidad del conjunto es deliberada: combinar familias de modelos con distintas formas funcionales y sesgos inductivos es precisamente lo que permite que sus errores no estén perfectamente correlacionados y que la dispersión entre ellos contenga información, en línea con el principio de diversificación de la combinación de pronósticos (Timmermann, 2006) y con la evidencia de que los modelos no lineales capturan en los retornos relaciones que las especificaciones lineales no recogen (Gu et al., 2020; Wang & Chen, 2023). El conjunto está compuesto por nueve especificaciones: un perceptrón multicapa (MLP), una red recurrente de memoria de largo y corto plazo (LSTM), dos variantes híbridas que incorporan volatilidad condicional, tres modelos de árboles y dos reglas de referencia basadas en promedios móviles. La Tabla 2 resume el conjunto, las referencias en que se apoya la definición de cada modelo y la forma en que cada uno representa la identidad del activo.

Todas las especificaciones comparten una misma formulación del problema y una mis-

ma disciplina temporal. La estimación se realiza bajo un esquema conjunto sobre el universo de activos: en lugar de ajustar un modelo independiente por acción, se construye una única muestra que agrupa las observaciones de todos los activos dentro de cada subperiodo, lo que amplía el tamaño muestral efectivo y permite explotar patrones comunes preservando, al mismo tiempo, la heterogeneidad transversal mediante una representación explícita del identificador del activo. Para cada activo i y fecha t se dispone de un vector de características, y el objetivo es predecir de forma simultánea el vector de retornos acumulados a los tres horizontes,

$$\widehat{Y}_{i,t} = \left(\widehat{R}_{i,t}^{(20)}, \widehat{R}_{i,t}^{(60)}, \widehat{R}_{i,t}^{(120)}\right)', \tag{3}$$

de modo que cada modelo entrenable resuelve una tarea de predicción multisalida y entrega un único valor escalar por horizonte. En los modelos que lo requieren, las variables se estandarizan utilizando estadísticos calculados exclusivamente sobre la muestra de entrenamiento, y los modelos entrenables minimizan el error cuadrático medio promediado sobre los tres horizontes. La forma de representar la identidad del activo distingue a las dos grandes familias: las redes neuronales incorporan un embedding entrenable, es decir, un vector de baja dimensión asociado a cada activo y aprendido conjuntamente con el resto de los parámetros, mientras que los modelos de árboles incorporan el identificador mediante codificación one-hot, lo que les permite aprender particiones específicas sin abandonar la muestra conjunta. Las dos reglas de promedio móvil no utilizan el identificador, ya que operan directamente sobre la serie de retornos de cada activo.

El MLP procesa la información contemporánea disponible en cada fecha y busca explotar relaciones no lineales entre las características y los retornos futuros. El embedding del activo se concatena con el vector de características al inicio del procesamiento, y la representación resultante atraviesa una secuencia de capas densas con regularización hasta una salida lineal de dimensión tres.

La LSTM (Hochreiter & Schmidhuber, 1997) incorpora explícitamente la trayectoria reciente de cada activo, procesando una secuencia de sesenta observaciones previas a la fecha de decisión con el fin de capturar persistencia, reversión y acumulación gradual de señales que una observación aislada no refleja. A diferencia del MLP, el embedding del activo no

Tabla 2: Composición del crowd predictivo: los nueve modelos, sus referencias y la representación de la identidad del activo. La última fila corresponde al algoritmo genético, que no integra el crowd sino que calibra los siete modelos entrenables.

Modelo	Referencia	Representación del activo
Perceptrón multicapa (MLP)	(Gu et al., 2020)	Embedding entrenable
Red recurrente LSTM	(Bhandari et al., 2022; Hochreiter & Schmidhuber, 1997)	Embedding entrenable
Híbrido GARCH-LSTM	(Bollerslev, 1986; García-Medina & Aguayo-Moreno, 2024; Hochreiter & Schmidhuber, 1997)	Embedding entrenable
Híbrido EGARCH-LSTM	(García-Medina & Aguayo-Moreno, 2024; Hochreiter & Schmidhuber, 1997; Nelson, 1991)	Embedding entrenable
Random Forest (RF)	(Breiman, 2001; Qian & Zhang, 2025)	Codificación one-hot
XGBoost	(T. Chen & Guestrin, 2016; Friedman, 2001)	Codificación one-hot
HistGradientBoosting (HGB)	(Friedman, 2001; Ke et al., 2017)	Codificación one-hot
Promedio móvil simple (MA)	—	No utiliza identificador
Promedio móvil exponencial (EMA)	—	No utiliza identificador
Algoritmo genético (GA)	(Alaminos et al., 2024; Thakkar & Chaudhari, 2024)	Optimiza los siete modelos entrenables

ingresa a la red recurrente, sino que se concatena con la representación secuencial una vez que esta ha resumido la trayectoria histórica. Esta decisión responde a la naturaleza distinta de ambas fuentes: el embedding es una representación estática del activo, idéntica a lo largo de toda la ventana, de modo que inyectarla en cada paso temporal obligaría a la red a propagar sin alteración una misma información, añadiendo parámetros sin ganancia predictiva. Procesar primero la dinámica temporal y combinarla después con la identidad transversal permite que cada componente del modelo se especialice en el tipo de información para el cual está diseñado. Esta arquitectura recurrente ha sido aplicada con éxito a la predicción de series financieras (Bhandari et al., 2022).

Las dos variantes híbridas comparten la estructura recurrente anterior y la amplían con una serie de volatilidad condicional incorporada como dimensión adicional de la secuencia de entrada, bajo la premisa de que la evolución esperada de los retornos depende también del régimen de riesgo del activo. La primera variante emplea la volatilidad estimada mediante un modelo de heterocedasticidad condicional autorregresiva generalizada (Bollerslev, 1986), que modela la varianza como función de las innovaciones recientes y de su propia persistencia; la segunda emplea una especificación exponencial (Nelson, 1991), que opera sobre el logaritmo

de la varianza y admite respuestas asimétricas ante shocks de distinto signo. En ambos casos las series de volatilidad se construyen de forma causal, mediante un esquema recursivo por bloques que reestima la especificación únicamente con retornos observados hasta cada fecha y, ante historia insuficiente, recurre a una desviación estándar móvil como respaldo. La integración de modelos de volatilidad condicional con redes recurrentes para la predicción financiera sigue la línea de trabajos híbridos recientes (García-Medina & Aguayo-Moreno, 2024).

Los tres modelos de árboles operan sobre la misma representación tabular y aportan capacidad para capturar interacciones no lineales y efectos umbral sin imponer una forma funcional. El RF (Breiman, 2001) pertenece a la familia de métodos de bagging, que promedian árboles construidos sobre remuestreos de la base y reducen así la varianza de la predicción agregada, y su utilidad para condensar relaciones no lineales en la predicción de retornos ha sido documentada en horizontes prolongados (Qian & Zhang, 2025). Las dos especificaciones de boosting, en cambio, construyen los árboles de forma secuencial, ajustando cada nuevo árbol a los errores del ensamble acumulado, lo que equivale a un descenso por gradiente en el espacio de funciones (Friedman, 2001). La primera de ellas, XGBoost (T. Chen & Guestrin, 2016), incorpora regularización explícita sobre la complejidad de los árboles junto con aleatorización sobre observaciones y variables; la segunda, HGB (Ke et al., 2017), discretiza previamente las variables continuas en intervalos, lo que reduce de forma sustantiva el costo de evaluar las particiones candidatas y la hace eficiente sobre bases de gran tamaño. La coexistencia de bagging y boosting, además de dos implementaciones distintas de este último, aumenta la diversidad efectiva del crowd y evita que la dispersión entre modelos refleje meras duplicaciones de una misma familia.

Finalmente, el conjunto incorpora dos reglas de referencia basadas en promedios móviles de los retornos diarios observados, una simple y una exponencial, que aproximan el retorno esperado diario y lo escalan al horizonte correspondiente. La longitud de la ventana, en el caso simple, y la memoria efectiva, en el exponencial, se fijan de forma determinística al horizonte de predicción. En el pipeline de retornos tratados, estas reglas reciben un filtrado causal consistente con el tratamiento del objetivo: sobre una ventana histórica de doscientos cuarenta retornos anclada en cada fecha de decisión se aplica la transformación wavelet seguida de

winsorización, estrictamente con información disponible hasta ese momento, y solo entonces se calcula el promedio. Estas referencias, parsimoniosas e interpretables, permiten evaluar si la complejidad adicional de los modelos no lineales y secuenciales se traduce efectivamente en mejoras predictivas fuera de muestra.

3.4.1. Optimización de hiperparámetros mediante algoritmo genético

Los siete modelos entrenables del crowd —el MLP, la LSTM, sus dos variantes híbridas y los tres modelos de árboles— contienen hiperparámetros que no se estiman minimizando la función de pérdida pero que condicionan de manera sustantiva su capacidad de generalización, como la profundidad de la arquitectura, el número de unidades o árboles, la tasa de aprendizaje y la intensidad de regularización. Dado que estos elementos afectan simultáneamente sesgo, varianza y estabilidad, su selección se realiza mediante un algoritmo genético (GA), procedimiento de optimización inspirado en la evolución natural cuyo uso para la calibración de modelos de predicción financiera se encuentra ampliamente documentado (Alaminos et al., 2024; Thakkar & Chaudhari, 2024). Las dos reglas de promedio móvil quedan excluidas de esta etapa, pues su única configuración —la longitud de ventana— se fija de forma determinística al horizonte.

El GA opera sobre una población de configuraciones candidatas, cada una representada como un vector cuyos componentes codifican los hiperparámetros de un modelo dentro de su espacio de búsqueda admisible. Cada configuración se evalúa según su desempeño en una muestra de validación temporal interna, construida reservando el tramo final del periodo de entrenamiento de forma que la selección de hiperparámetros nunca utilice información posterior y se replique la lógica fuera de muestra sin tocar el tramo de prueba. La población evoluciona a lo largo de sucesivas generaciones mediante selección por torneo, en la que se escoge la mejor de un subconjunto aleatorio de configuraciones; cruzamiento, que combina segmentos de dos configuraciones; mutación, que reemplaza algunos componentes por valores admisibles; y elitismo, que preserva la mejor configuración de cada generación para que no se pierda por efecto de la recombinación. En las redes neuronales, la evaluación de cada configuración incorpora mecanismos de estabilización propios del entrenamiento neuronal, como la detención temprana

con restauración de los mejores pesos y la reducción adaptativa de la tasa de aprendizaje, y se registra además la época de mejor desempeño para el reentrenamiento final. La configuración general del procedimiento se presenta en la Tabla 3 y los espacios de búsqueda por familia de modelos en la Tabla 4.

Tabla 3: Configuración general del algoritmo genético.

Parámetro	Valor
Tamaño de población	10
Número de generaciones	8
Probabilidad de mutación	0,20
Probabilidad de cruzamiento	0,90
Tamaño del torneo	2
Elitismo	1 individuo
Semilla de reproducibilidad	123
Fracción de validación temporal interna	0,15
Criterio de selección	Minimización de la pérdida de validación

Tabla 4: Espacios de búsqueda de hiperparámetros por familia de modelos.

Familia de modelos	Hiperparámetro	Tipo	Rango de búsqueda
MLP	Número de capas ocultas	Entero	1–2
	Unidades de la primera capa	Entero	12–256
	Unidades de la segunda capa	Entero	12–128
	Tasa de dropout	Continuo	0,10–0,35
	Tasa de aprendizaje	Continuo	1×10^{-4}–5×10^{-3}
	Tamaño de lote	Entero	64–512
	Épocas	Entero	12–120
	Dimensión del embedding	Entero	4–16
LSTM y variantes híbridas	Número de capas recurrentes	Entero	1–2
	Unidades de la primera capa	Entero	12–256
	Unidades de la segunda capa	Entero	12–256
	Unidades de la capa densa	Entero	16–96
	Tasa de dropout	Continuo	0,10–0,35
	Tasa de aprendizaje	Continuo	1×10^{-4}–5×10^{-3}
	Tamaño de lote	Entero	64–256
	Épocas	Entero	12–120
	Dimensión del embedding	Entero	4–16
RF	Número de árboles	Entero	150–500
	Profundidad máxima	Entero	4–16
	Mínimo de observaciones por hoja	Entero	1–15
XGBoost	Número de árboles	Entero	150–500
	Profundidad máxima	Entero	3–10
	Tasa de aprendizaje	Continuo	0,01–0,20
	Submuestreo de observaciones	Continuo	0,70–1,00
	Submuestreo de variables	Continuo	0,70–1,00
	Penalización L_1	Continuo	0,00–1,00
	Penalización L_2	Continuo	0,10–3,00
HGB	Número máximo de iteraciones	Entero	100–400
	Tasa de aprendizaje	Continuo	0,01–0,20
	Profundidad máxima	Entero	3–12
	Mínimo de observaciones por hoja	Entero	5–40

Aunque la Tabla 4 agrupa los hiperparámetros por familia, su función puede caracterizarse de forma transversal en cuatro categorías. Un primer grupo gobierna la complejidad estructural del modelo: el número de capas y de unidades en las redes neuronales, y el tamaño

del ensamble y la profundidad de los árboles en los modelos basados en árboles. Un segundo grupo regula la intensidad de regularización: la tasa de dropout en las redes, el tamaño mínimo de las hojas terminales en los modelos de árboles y las penalizaciones sobre los pesos de las hojas en el boosting. Un tercer grupo controla el proceso de optimización, e incluye la tasa de aprendizaje, el tamaño de lote, el número de épocas y, en una de las especificaciones de boosting, los mecanismos de aleatorización sobre observaciones y variables. Un cuarto componente, exclusivo de las redes neuronales, corresponde a la dimensión del embedding del activo, que regula la capacidad del modelo para representar la heterogeneidad transversal del universo. De esta manera, la diversidad del crowd no proviene únicamente de combinar familias algorítmicas distintas, sino también de permitir que cada familia adopte una configuración interna ajustada a su desempeño fuera de muestra dentro del tramo de entrenamiento.

3.5 Entrenamiento, predicción fuera de muestra y métricas predictivas

Una vez seleccionada la configuración de hiperparámetros mediante el algoritmo genético, cada modelo entrenable se reentrena utilizando la totalidad de la muestra de entrenamiento del subperiodo correspondiente. La distinción entre ambas etapas es deliberada: la validación temporal interna sirve únicamente para seleccionar configuraciones, mientras que el ajuste final aprovecha toda la información disponible antes del inicio del periodo de evaluación. En las redes neuronales, este reentrenamiento emplea el número de épocas asociado al punto de mejor generalización registrado durante la búsqueda, de modo que la duración del ajuste final sea coherente con el desempeño de validación observado. En ningún caso el ajuste final incorpora observaciones pertenecientes al tramo de prueba.

La disciplina temporal se preserva también en la aplicación de las transformaciones a los datos. Los estadísticos de estandarización de las redes neuronales se estiman exclusivamente sobre la muestra de entrenamiento y se aplican sin modificación al tramo de evaluación, evitando cualquier filtración de información desde el periodo de prueba. En los modelos secuenciales, las primeras ventanas del tramo de evaluación se construyen incorporando el contexto histórico inmediatamente anterior, de manera que la formación de cada secuencia no requiera observaciones futuras. La fuente de la variable objetivo durante el entrenamiento depende del pipeline

considerado, según se describió previamente, pero la evaluación fuera de muestra se realiza en todos los casos contra retornos realizados en escala cruda, lo que asegura la comparabilidad económica entre ambos pipelines.

Las predicciones fuera de muestra no se generan para todas las fechas del periodo de evaluación, sino sobre un calendario de rebalanceo construido de forma estructurada. Para cada subperiodo se define una ventana de doscientos cuarenta días hábiles al inicio del tramo de prueba. Dentro de esa ventana, y para cada horizonte $h \in \{20, 60, 120\}$, se construye un calendario de rebalanceo tomando fechas espaciadas cada h observaciones a partir del inicio de la ventana, lo que genera doce fechas de decisión para el horizonte de veinte días, cuatro para el de sesenta y dos para el de ciento veinte. El conjunto de fechas ancla se define entonces como la unión de los calendarios de los tres horizontes,

$$\mathscr{T}^A = \mathscr{T}^{(20)} \cup \mathscr{T}^{(60)} \cup \mathscr{T}^{(120)}. \tag{4}$$

Cada modelo genera predicciones únicamente para las fechas pertenecientes a este conjunto, y posteriormente, para cada horizonte, se conservan solo las predicciones asociadas a su calendario de rebalanceo específico. Esta organización cumple un doble propósito: permite almacenar de forma compacta una matriz de predicciones indexada por fecha ancla y, al mismo tiempo, construir un panel en formato largo directamente utilizable en las etapas posteriores de consenso, desacuerdo y construcción de portafolios, manteniendo en todo momento la coherencia entre la frecuencia de decisión y el horizonte de cada predicción.

Cada predicción se interpreta como una aproximación al retorno logarítmico acumulado esperado del activo en el horizonte correspondiente, condicional a la información disponible en la fecha de decisión. Denotando por $\widehat{R}^{(h)}_{i,t,m}$ la predicción generada por el modelo m para el activo i, la fecha t y el horizonte h, esta puede expresarse como

$$\widehat{R}^{(h)}_{i,t,m} \approx \mathbb{E}^m_t \left[\sum_{k=1}^{h} r_{i,t+k} \right], \tag{5}$$

donde $\mathbb{E}^m_t[\cdot]$ denota la esperanza condicional al conjunto de información disponible en t bajo la representación implícita aprendida por el modelo m. En términos operativos, cada modelo en-

trega un único valor escalar por horizonte que aproxima directamente esa esperanza acumulada, sin reconstruir explícitamente los retornos diarios intermedios. Estas predicciones constituyen las opiniones individuales del crowd que alimentan las etapas posteriores del marco.

La calidad predictiva de cada modelo se evalúa comparando sus predicciones fuera de muestra con los retornos realizados correspondientes, construidos siempre a partir de retornos crudos observados, para cada combinación de subperiodo, pipeline, modelo y horizonte. Se consideran tres métricas complementarias. La raíz del error cuadrático medio (RMSE) resume la precisión global penalizando con mayor intensidad los errores grandes; el error absoluto medio (MAE) ofrece una medida de error promedio menos sensible a observaciones extremas; y la precisión de signo (ACC) evalúa la capacidad direccional del modelo, es decir, su acierto en anticipar el signo del retorno acumulado futuro. Mientras las dos primeras cuantifican la magnitud del error, la última captura una dimensión distinta y especialmente relevante para la asignación de cartera, en la medida en que el ordenamiento de los activos depende más del signo y la posición relativa de las predicciones que de su magnitud exacta. El conjunto de predicciones fuera de muestra y sus métricas asociadas constituye la base empírica del crowd artificial: las etapas siguientes no reestiman los modelos, sino que toman estas predicciones como señales heterogéneas, comparables y cronológicamente disciplinadas.

3.6 Construcción y homogeneización de las señales del crowd

Una vez generadas las predicciones fuera de muestra, estas se reorganizan en una estructura apta para estudiar el acuerdo y la discrepancia entre fuentes de información. El objetivo de esta etapa ya no es producir nuevos pronósticos, sino transformar el conjunto de predicciones individuales en un sistema de señales agregables y comparables. Para cada combinación de subperiodo, pipeline y horizonte, las predicciones se disponen en un panel en formato largo indexado por activo, fecha y modelo (i,t,m), donde i identifica el activo, t la fecha de decisión y m el modelo emisor. Esta representación hace explícito el problema de agregación de opiniones: para un mismo activo y fecha coexisten múltiples señales emitidas por metodologías distintas, que posteriormente pueden resumirse mediante un operador de consenso o caracterizarse mediante una medida de dispersión. La heterogeneidad entre modelos, inducida de antemano

por la diversidad del crowd, constituye una condición necesaria para que ambas magnitudes resulten informativas: si todas las fuentes coincidieran, no existiría desacuerdo que medir ni diversidad de perspectivas que aprovechar.

Dado que las predicciones de modelos heterogéneos pueden diferir en escala, dispersión y presencia de valores extremos, su agregación directa podría reflejar discrepancias puramente mecánicas asociadas al nivel o la variabilidad de cada modelo, y no diferencias informacionales genuinas. Para evitarlo, las señales se someten a una homogeneización transversal que opera en dos pasos —una winsorización seguida de una estandarización—, ambos aplicados de forma independiente para cada fecha y modelo sobre la sección cruzada de activos.

El primer paso aplica una winsorización transversal que acota las predicciones por debajo del percentil uno por ciento y por encima del percentil noventa y nueve por ciento de la sección cruzada de cada modelo en cada fecha, atenuando la influencia de valores extremos. Conviene subrayar que esta winsorización transversal es distinta de la winsorización temporal aplicada a la variable objetivo durante el entrenamiento (Sección 3.3): aquella actuaba sobre la serie de retornos de cada activo dentro del tramo de entrenamiento, mientras que esta opera sobre la sección cruzada de predicciones de un mismo modelo en una fecha determinada. La predicción winsorizada resultante, que denotamos $\widetilde{R}^{(h)}_{i,t,m}$ para el modelo m, el activo i, la fecha t y el horizonte h, no constituye una representación empleada de forma autónoma en las etapas posteriores, sino un insumo intermedio que sirve exclusivamente como entrada del segundo paso.

El segundo paso construye sobre la serie winsorizada una estandarización transversal que expresa las predicciones de cada modelo en unidades comparables,

$$x^{(h)}_{i,t,m} = \frac{\widetilde{R}^{(h)}_{i,t,m} - \mu_t^{(h,m)}}{\sigma_t^{(h,m)}}, \tag{6}$$

donde $\mu_t^{(h,m)}$ y $\sigma_t^{(h,m)}$ corresponden, respectivamente, a la media y la desviación estándar transversal sobre los N_t activos de la sección cruzada de las predicciones winsorizadas del modelo m para el horizonte h en la fecha t, esta última calculada con $N_t - 1$ grados de libertad. En el caso degenerado en que la desviación estándar transversal sea numéricamente nula, la señal estandarizada se fija en cero para toda la sección cruzada de ese par modelo-fecha. Esta estan-

darización centra y escala las señales utilizando únicamente información contemporánea de la sección cruzada, de modo que un modelo con mayor varianza predictiva no domine mecánicamente las medidas agregadas sin aportar por ello mayor contenido informacional. Una vez obtenida la señal estandarizada, las predicciones expresadas en su escala original dejan de utilizarse, y $x^{(h)}_{i,t,m}$ —adimensional y comparable entre modelos— se erige como la representación operativa única de cada predicción individual.

Esta representación común constituye el insumo a partir del cual se construyen las dos magnitudes centrales del marco. Sobre ella se define el consenso del crowd, que aproxima el componente direccional de la señal informacional agregada y que posteriormente proveerá el vector de retornos esperados del problema de cartera; y sobre ella se define el desacuerdo entre modelos, que aproxima el componente de riesgo informacional y que proveerá la matriz de riesgo. Utilizar una misma representación homogeneizada en ambas dimensiones asegura coherencia dimensional entre los insumos que alimentan la asignación de portafolio, y constituye una pieza estructural de la formulación informacional adoptada en este trabajo. La construcción explícita del consenso y del desacuerdo a partir de esta señal se desarrolla en la subsección siguiente.

3.7 Construcción estática de consenso y desacuerdo

A partir de la señal homogeneizada se construyen las magnitudes agregadas que sintetizan la opinión del crowd para cada activo y fecha. En esta subsección se definen bajo esquemas de agregación estáticos, que cumplen un doble propósito: sirven como insumo para la construcción de portafolios y, a la vez, como punto de comparación frente a los esquemas adaptativos introducidos más adelante. Todas las magnitudes que siguen se construyen de forma independiente para cada subperiodo, pipeline y horizonte, y toman como insumo común la señal estandarizada $x^{(h)}_{i,t,m}$ generada en la etapa de homogeneización, lo que asegura coherencia dimensional entre el componente direccional y el componente de dispersión de la opinión colectiva.

Sea $\mathcal{M}_t$ el conjunto de modelos disponibles en la fecha t. El consenso del crowd resume la dirección de la opinión colectiva sobre cada activo y se obtiene agregando las señales

individuales sobre el conjunto de modelos. La medida base es el promedio entre modelos,

$$C_{i,t}^{(h),\text{mean}} = \frac{1}{|\mathscr{M}_t|} \sum_{m \in \mathscr{M}_t} x_{i,t,m}^{(h)}, \tag{7}$$

mientras que una alternativa robusta frente a opiniones extremas se obtiene mediante la mediana,

$$C_{i,t}^{(h),\text{med}} = \text{mediana}_{m \in \mathscr{M}_t} \left(x_{i,t,m}^{(h)} \right). \tag{8}$$

Ambas expresan la posición relativa de cada activo respecto del resto del mercado en una escala común y adimensional, de modo que el consenso debe interpretarse como alineamiento direccional relativo entre activos antes que como magnitud absoluta de retorno esperado.

El desacuerdo entre modelos mide la dispersión de las opiniones individuales en torno al consenso y constituye el componente de riesgo informacional del marco. La literatura de combinación de pronósticos reconoce precisamente la dispersión entre las predicciones individuales como una medida de la incertidumbre que rodea al consenso (Timmermann, 2006). La medida base es la desviación estándar entre modelos,

$$D_{i,t}^{(h),\text{std}} = \sqrt{\frac{1}{|\mathscr{M}_t| - 1} \sum_{m \in \mathscr{M}_t} \left(x_{i,t,m}^{(h)} - C_{i,t}^{(h),\text{mean}} \right)^2}, \tag{9}$$

y una versión robusta se obtiene mediante el rango intercuartílico,

$$D_{i,t}^{(h),\text{iqr}} = Q_{0,75} \left(x_{i,t,m}^{(h)} \right) - Q_{0,25} \left(x_{i,t,m}^{(h)} \right), \tag{10}$$

donde los cuantiles se calculan sobre el conjunto de modelos $\mathscr{M}_t$. Un mayor desacuerdo indica que los modelos del crowd sostienen opiniones más dispares sobre la posición relativa del activo, lo que se interpreta como mayor incertidumbre informacional respecto de la calidad de la señal de consenso, en línea con la lectura del desacuerdo como dimensión de riesgo desarrollada en la Sección 2 (Carlin et al., 2014; Diether et al., 2002).

La interacción entre la fuerza de la opinión colectiva y su dispersión interna permite construir una medida sintética de riesgo informacional relativo. La intuición es que una señal

resulta más sólida cuanto mayor es su magnitud direccional y menor su nivel de desacuerdo. Esta idea se operacionaliza mediante la convicción inversa,

$$D_{i,t}^{(h),\text{invconv}} = \frac{D_{i,t}^{(h),\text{std}} + \varepsilon}{\left| C_{i,t}^{(h),\text{mean}} \right| + \varepsilon}, \tag{11}$$

donde $\varepsilon > 0$ es una constante pequeña que evita divisiones por cero. Valores elevados identifican activos cuya señal agregada es débil en relación con la dispersión interna del crowd, y por tanto de mayor riesgo informacional. Al construirse íntegramente sobre la señal homogeneizada, numerador y denominador comparten la misma escala adimensional, de modo que el cociente resulta comparable entre activos y fechas.

Finalmente, el desacuerdo admite un resumen a nivel de mercado mediante un indicador que promedia transversalmente, sobre los activos disponibles en cada fecha, la dispersión individual entre modelos,

$$D_{\text{MKT},t}^{(h),\text{EW}} = \frac{1}{N_t} \sum_{i=1}^{N_t} D_{i,t}^{(h),\text{std}}. \tag{12}$$

Este indicador caracteriza el entorno informacional agregado del mercado en cada fecha, resumiendo el nivel global de discrepancia del crowd. Como se detalla en la subsección siguiente, esta caracterización del régimen informacional se incorpora, junto con otras variables de entorno y de estructura del crowd, al vector de estado sobre el que opera el agente adaptativo.

En conjunto, estas magnitudes conforman el repertorio de señales informacionales del crowd que alimenta tanto los esquemas estáticos de construcción de portafolios como las extensiones adaptativas. Los consensos proporcionan el componente direccional de la asignación, los desacuerdos el componente de riesgo informacional, la convicción inversa una medida sintética que combina ambos en una única magnitud adimensional, y el indicador de mercado una caracterización del régimen informacional agregado. La forma en que estas señales se transforman en decisiones de cartera, primero bajo agregación estática y luego bajo agregación adaptativa, se desarrolla en las subsecciones siguientes.

3.8 Consenso adaptativo mediante aprendizaje secuencial

3.8.1. Motivación y formulación como problema de decisión secuencial

Cuando varios modelos predictivos están disponibles de forma simultánea, cada uno puede interpretarse como una fuente de creencias sobre los retornos futuros, heterogénea por construcción al provenir de estructuras estadísticas, algoritmos y esquemas de regularización distintos. La forma más simple de combinar esas opiniones consiste en aplicar reglas de agregación estáticas, como la ponderación equitativa o un resumen robusto fijo, que asumen de manera implícita que la relación entre los pronósticos individuales y la señal agregada permanece estable en el tiempo. En entornos financieros con quiebres estructurales, regímenes cambiantes de volatilidad y rotación en el liderazgo transversal de los activos, ese supuesto resulta excesivamente fuerte: un modelo valioso en fases de tendencia persistente puede perder relevancia en episodios de reversión, y especificaciones orientadas a relaciones no lineales o a componentes de baja frecuencia pueden dominar solo en ciertos regímenes u horizontes. La regla de agregación óptima debería, por tanto, poder variar a través del tiempo, en línea con la necesidad de adaptación que imponen la no estacionariedad y los cambios de régimen de las series financieras (Shahparast et al., 2023).

Esta observación motiva el tránsito desde la combinación estática hacia una agregación secuencial adaptativa. En lugar de asignar pesos permanentes a los integrantes del crowd, el objetivo es determinar, en cada fecha de rebalanceo, qué combinación de modelos resulta más conveniente dado el conjunto de información disponible. La lógica es análoga a un problema de delegación bajo incertidumbre: un decisor que recibe opiniones de múltiples fuentes con especializaciones y trayectorias distintas reasigna su atención hacia aquellas actualmente más informativas, preservando diversificación cuando la incertidumbre es elevada. Esta reasignación afecta simultáneamente a las dos magnitudes centrales del marco, pues modifica tanto el consenso contenido en el pronóstico combinado como el grado efectivo de desacuerdo entre las fuentes más ponderadas. Aprender a ponderar el crowd equivale, así, a aprender a extraer de manera conjunta información direccional y riesgo informacional a partir de expectativas heterogéneas.

Formalmente, el problema de agregación consiste en escoger en cada fecha t un vector de pesos sobre los $M = 9$ modelos del crowd,

$$w_t = (w_{1,t}, \ldots, w_{M,t}), \qquad w_{m,t} \geq 0, \qquad \sum_{m=1}^{M} w_{m,t} = 1, \tag{13}$$

de modo que la combinación resultante refleje qué modelos son relativamente más informativos en esa fecha. Estos pesos no son constantes predeterminadas, sino el resultado endógeno de un mecanismo de aprendizaje que actualiza sus decisiones a medida que se incorpora nueva información, y se restringen a un conjunto finito de combinaciones admisibles sobre el simplex, detallado en la subsubsección de espacio de acciones (Sección 3.8.3).

La estructura del problema es naturalmente secuencial y se formula como un proceso de decisión de Markov. En cada fecha de rebalanceo t, el agente observa un estado S_t que resume el entorno económico, la estructura interna del crowd y el desempeño reciente de los modelos; selecciona una acción a_t que indexa un vector de pesos $w(a_t)$ sobre el simplex de modelos; y, tras avanzar a la siguiente fecha de decisión, recibe una recompensa escalar $\mathscr{R}_t$ definida a partir del desempeño realizado del portafolio inducido por esa decisión y mantenido durante el horizonte correspondiente. De esta manera, una combinación de pronósticos resulta valiosa en la medida en que mejora los resultados de inversión posteriores, y no simplemente por reducir errores predictivos contemporáneos, lo que introduce una disyuntiva intertemporal ausente en los enfoques supervisados tradicionales.

El objetivo del agente es aprender una política $\pi : \mathscr{S} \to \mathscr{A}$ que asigne una acción conveniente a cada estado. La calidad de una política se resume mediante la función de valor-acción,

$$Q^{\pi}(s,a) = \mathbb{E}\left[\sum_{k=0}^{K_t} \gamma^k \mathscr{R}_{t+k} \,\middle|\, S_t = s,\ a_t = a,\ \pi\right], \tag{14}$$

donde $\gamma \in [0,1)$ es el factor de descuento intertemporal y K_t el número de fechas de decisión restantes dentro del episodio iniciado en t. Cada episodio queda delimitado por una combinación de subperiodo y horizonte de mantención, de modo que su última transición se trata como terminal y trunca el descuento. La implementación de esta formulación se desarrolla mediante una arquitectura Double Deep Q-Network, adecuada para problemas con estados de alta di-

mensión, recompensas ruidosas y decisiones discretas, cuya descripción requiere precisar antes la construcción del vector de estado y del espacio de acciones.

3.8.2. Representación del estado

El vector de estado resume la información disponible en cada fecha de rebalanceo y condiciona directamente las decisiones del agente. Debe contener señales suficientes para distinguir entre regímenes de mercado, cambios en la estructura interna del crowd, variaciones en el desempeño relativo de los modelos y diferencias asociadas al horizonte de inversión. Con ese fin, en cada fecha t se define un vector $S_t \in \mathbb{R}^{37}$ compuesto por cuatro bloques complementarios,

$$S_t = \left(Z_t^{\text{mkt}}, Z_t^{\text{crowd}}, Z_t^{\text{perf}}, Z_t^{\text{meta}} \right), \tag{15}$$

correspondientes a información de mercado, estructura del crowd, desempeño histórico de los modelos y contexto operativo.

El bloque de mercado, $Z_t^{\text{mkt}} = (D_{\text{MKT},t}^{(h),\text{EW}}, \text{RET}_{h,t}, \text{VOL}_{h,t}, \text{DD}_{h,t})$, reúne cuatro variables. La primera es el índice de desacuerdo de mercado equiponderado definido en la subsección anterior, que caracteriza el entorno informacional agregado. Las tres restantes resumen la dinámica reciente del benchmark sobre una ventana móvil de los h días bursátiles previos a la fecha de decisión, donde h coincide con el horizonte de mantención del agente: el log-retorno acumulado del benchmark en esa ventana, la desviación estándar de sus log-retornos diarios y el máximo drawdown de su trayectoria acumulada. En conjunto, este bloque permite distinguir entre períodos expansivos, episodios de estrés y entornos de mayor incertidumbre.

El bloque de crowd, $Z_t^{\text{crowd}} = (\rho_t^{\text{mean}}, \text{PC1}_t^{\text{share}})$, resume la estructura interna de las opiniones. Organizando la sección cruzada de señales homogeneizadas en una matriz activo-modelo y calculando su matriz de correlación entre modelos $\mathbf{C}_t$, de elementos $c_{m\ell,t}$, la primera variable es el promedio de las correlaciones fuera de la diagonal,

$$\rho_t^{\text{mean}} = \frac{2}{M(M-1)} \sum_{m<\ell} c_{m\ell,t}, \tag{16}$$

cuyos valores altos indican opiniones fuertemente alineadas y los bajos mayor heterogeneidad.

La segunda se obtiene de un análisis de componentes principales sobre la misma matriz: si $\lambda_{1,t} \geq \cdots \geq \lambda_{K_t,t}$ son sus autovalores ordenados, con $K_t = \text{mín}(N_t, M)$, la fracción de varianza explicada por el primer componente es

$$\text{PC1}_t^{\text{share}} = \frac{\lambda_{1,t}}{\sum_{k=1}^{K_t} \lambda_{k,t}}. \tag{17}$$

Una participación elevada señala que la variación conjunta del crowd se resume en un único factor común, indicando una estructura concentrada, mientras que valores bajos reflejan información distribuida y un crowd más diverso.

El bloque de desempeño incorpora la calidad predictiva reciente de cada modelo. Para los $M = 9$ modelos se calculan tres métricas acumuladas hasta la fecha de decisión —la raíz del error cuadrático medio, el error absoluto medio y la tasa de acierto direccional, definidas en la Sección 3.5—, que se apilan consecutivamente por modelo según un orden fijo para conformar un vector Z_t^{perf} de veintisiete componentes. Estas estadísticas se actualizan bajo un esquema expandido, recalculándose en cada fecha con toda la información observada desde el inicio del tramo hasta la fecha previa. Dado que en la primera fecha de rebalanceo aún no existe historia interna suficiente, su inicialización emplea un prior construido exclusivamente con información de la ventana de entrenamiento previa del crowd, lo que evita estados vacíos, preserva la disciplina temporal y mantiene coherencia entre entrenamiento y evaluación; desde la segunda fecha en adelante, las métricas se actualizan solo con información del nuevo período.

Finalmente, el bloque meta, $Z_t^{\text{meta}} = (\text{STEP}_t, H20_t, H60_t, H120_t)$, incorpora el contexto operativo. La variable STEP_t representa la posición temporal normalizada dentro del período de decisión mediante una tendencia lineal creciente entre cero y uno, que permite distinguir si una misma configuración del entorno ocurre en una etapa temprana o avanzada del ciclo. Las tres variables restantes son indicadores binarios mutuamente excluyentes del horizonte de mantención, que toman valor uno para el horizonte de veinte, sesenta o ciento veinte días, respectivamente, reconociendo que una misma configuración de mercado puede justificar ponderaciones distintas según el horizonte sobre el que se evaluará el portafolio. La dimensión total

satisface

$$\dim(S_t) = 37 = \underbrace{4}_{Z_t^{\text{mkt}}} + \underbrace{2}_{Z_t^{\text{crowd}}} + \underbrace{27}_{Z_t^{\text{perf}}} + \underbrace{4}_{Z_t^{\text{meta}}}. \tag{18}$$

Previo al entrenamiento, los componentes del estado se escalan mediante transformaciones estandarizadas estimadas únicamente sobre la muestra de entrenamiento del meta-agente, lo que homogeneiza las magnitudes entre bloques y favorece la estabilidad numérica sin incorporar información fuera de muestra.

3.8.3. Espacio de acciones

La política del agente actúa sobre un conjunto finito y predefinido de vectores de pesos. Esta elección busca un balance entre flexibilidad y manejabilidad: un espacio continuo sobre el simplex es teóricamente más rico, pero exige arquitecturas considerablemente más complejas y resulta más sensible al ruido de la recompensa, especialmente bajo episodios cortos. La discretización permite emplear arquitecturas de valor-acción estándar y reduce la varianza del aprendizaje, a costa de restringir las decisiones a un conjunto finito de combinaciones. Formalmente, el espacio de acciones es un conjunto $\mathscr{A} = \{w^{(1)}, \ldots, w^{(N)}\}$ de cardinalidad $N = 50$, donde cada elemento pertenece al simplex de probabilidades sobre los $M = 9$ modelos. El vocabulario se construye combinando cinco bloques diseñados para cubrir distintos grados de concentración y diversificación.

El primer bloque es una única acción de ponderación equitativa, $w^{\text{eq}} = \frac{1}{M}\mathbf{1}_M$, que asigna el mismo peso a todos los modelos y representa la ausencia de discriminación informacional entre fuentes. El segundo bloque se compone de cuatro acciones concentradas, una por cada familia metodológica, representada por un modelo prototípico: MLP para redes neuronales, GARCH-LSTM para híbridos, RF para árboles y EMA para los modelos de referencia. Cada una asigna peso unitario a su representante,

$$w^{(f)} = \mathbf{e}_{r_f}, \qquad f \in \{\text{NN}, \text{HYBRID}, \text{TREES}, \text{BASELINE}\}, \tag{19}$$

donde $\mathbf{e}_m$ es el vector canónico que concentra el peso en el modelo m, permitiendo al agente delegar la decisión en una sola familia cuando el contexto lo justifique.

El tercer bloque introduce dieciocho combinaciones equitativas top-k, definidas a partir de subconjuntos $\mathscr{S} \subset \{1,\ldots,M\}$ de tamaños $k \in \{2,3,4\}$ que asignan peso uniforme $1/k$ a los modelos seleccionados,

$$w_m^{(\mathscr{S})} = \begin{cases} 1/k & \text{si } m \in \mathscr{S}, \\ 0 & \text{en otro caso,} \end{cases} \tag{20}$$

incluyendo seis subconjuntos por cada valor de k y capturando formas intermedias de concentración entre los singletons y la ponderación equitativa. El cuarto y el quinto bloque corresponden a mezclas muestreadas de una distribución de Dirichlet, $w^{(j)} \sim \mathrm{Dir}(\alpha \mathbf{1}_M)$: quince realizaciones dispersas con parámetro de concentración bajo, $\alpha = 0{,}3$, que inducen distribuciones desbalanceadas en las que pocos modelos concentran la mayor parte del peso, y doce realizaciones densas con parámetro elevado, $\alpha = 2{,}0$, que se aproximan a la ponderación equitativa introduciendo perturbaciones suaves que rompen la simetría.

Los cinco bloques suman exactamente $1+4+18+15+12 = 50$ vectores. Tras su construcción, todos se renormalizan dividiendo por la suma de sus componentes, de modo que la pertenencia al simplex se mantiene exacta hasta precisión numérica. Los componentes estocásticos del vocabulario, esto es, los subconjuntos top-k y las realizaciones de Dirichlet, se generan con una semilla fija, por lo que el conjunto $\mathscr{A}$ es determinista y reproducible entre ejecuciones. Esta estructura cubre un espectro amplio de comportamientos, desde la máxima concentración de los singletons, pasando por las ponderaciones intermedias de los bloques top-k y Dirichlet disperso, hasta los esquemas casi equitativos del Dirichlet denso y la ponderación uniforme. El agente no aprende a producir pesos arbitrarios, sino a seleccionar cuál de estas configuraciones predefinidas resulta más adecuada en cada estado observado.

3.8.4. Función de recompensa

Definidos el estado y el espacio de acciones, el componente que completa el proceso de decisión es el criterio con que el agente evalúa sus elecciones. En aplicaciones financieras este criterio es determinante, pues una función objetivo mal especificada puede inducir políticas con buen ajuste estadístico pero escaso valor económico una vez implementadas. Por esa razón, la recompensa se define directamente en términos del desempeño del portafolio inducido, y no a

partir de métricas predictivas intermedias.

En cada fecha de rebalanceo, la acción seleccionada determina un vector de pesos sobre los modelos del crowd, a partir del cual se construyen el consenso y el desacuerdo adaptativos para los activos disponibles. Ambas señales alimentan el problema de máximo ratio de Sharpe informacional descrito en la Sección 3.9 —el consenso adaptativo como componente direccional y el desacuerdo adaptativo como matriz diagonal de riesgo informacional, ambos en la escala estandarizada homogeneizada—, lo que genera un portafolio implementable que se mantiene durante el horizonte $h \in \{20, 60, 120\}$. La recompensa de la decisión adoptada en t se define como el ratio de Sharpe realizado de ese portafolio durante la ventana de mantención. Formalmente,

$$\mathscr{R}_t = \frac{\frac{1}{h}\sum_{k=1}^{h}\left(r_{p,t+k} - r_{f,t+k}\right)}{\sqrt{\frac{1}{h-1}\sum_{k=1}^{h}\left[\left(r_{p,t+k} - r_{f,t+k}\right) - \frac{1}{h}\sum_{j=1}^{h}\left(r_{p,t+j} - r_{f,t+j}\right)\right]^2}}.$$

Aunque la recompensa se indexa en t, su valor utiliza información observada entre $t+1$ y $t+h$, ya que evalúa de forma posterior el desempeño de la decisión implementada. En esta expresión, $r_{p,t+k}$ representa el retorno realizado del portafolio en el día $t+k$ y $r_{f,t+k}$ la tasa libre de riesgo correspondiente al mismo día. El numerador captura el exceso de retorno promedio generado por la cartera durante la ventana de inversión, mientras que el denominador mide la volatilidad realizada de dicho exceso de retorno. En consecuencia, la recompensa aumenta cuando la decisión tomada en t induce portafolios que combinan mayor rentabilidad con menor variabilidad durante el horizonte de mantención.

El uso de un ratio de Sharpe como recompensa incorpora de manera explícita la disyuntiva entre retorno y riesgo: una función centrada únicamente en retorno incentivaría políticas excesivamente agresivas, mientras que una basada solo en error predictivo ignoraría la estructura de riesgo de la cartera resultante. Así, el agente aprende a valorar las combinaciones de modelos no solo por su capacidad de anticipar movimientos favorables, sino por la estabilidad económica de las carteras que inducen, alineando el aprendizaje con el objetivo final del inversionista. Esta definición reconoce, además, que distintos horizontes de mantención generan recompensas en escalas diferentes, pues las ventanas más extensas tienden a producir volatili-

dades realizadas de magnitud distinta a las más cortas. Esta heterogeneidad de escala motiva un tratamiento de las recompensas previo a su uso en el algoritmo de aprendizaje, que se detalla en la Sección 3.8.6.

3.8.5. Arquitectura Double Deep Q-Network

Para aprender la política, la función de valor-acción se aproxima mediante una red neuronal paramétrica $Q_\theta(S_t, a)$ que, dado un estado, entrega una estimación del valor esperado de cada una de las $N = 50$ acciones del vocabulario. La política selecciona la acción de mayor valor estimado,

$$\pi_\theta(S_t) = \arg\max_{a \in \mathcal{A}} Q_\theta(S_t, a), \tag{21}$$

regla que durante el entrenamiento se complementa con un mecanismo de exploración ε-greedy descrito en la sección siguiente. La red es completamente conectada, con una capa de entrada de dimensión treinta y siete, dos capas ocultas de sesenta y cuatro y treinta y dos unidades con activación ReLU y dropout de tasa 0,10, y una capa de salida lineal de dimensión cincuenta, en la que cada nodo corresponde al valor estimado de una acción. Esta capacidad acotada permite capturar interacciones no lineales entre las variables de mercado, las métricas del crowd y el desempeño histórico de los modelos, reduciendo a la vez el riesgo de sobreajuste sobre series financieras de tamaño muestral limitado.

El Q-learning tradicional adolece de una sobreestimación sistemática de los valores cuando una misma red se utiliza para seleccionar y evaluar las acciones futuras. Para mitigarlo, se adopta una arquitectura Double DQN que, sobre la base del Deep Q-Network —el cual estabiliza el entrenamiento mediante experience replay y una red objetivo separada (Mnih et al., 2015)—, desacopla la selección de la evaluación de la acción (van Hasselt et al., 2016). La implementación emplea dos redes de idéntica topología y parámetros desacoplados: una red online Q_θ, que selecciona acciones y se actualiza por descenso de gradiente, y una red objetivo Q_{θ^-}, que construye objetivos temporales más estables y cuyos parámetros siguen a los de la red online de manera diferida.

El entrenamiento se apoya en la ecuación de Bellman, que en la variante Double DQN

define el objetivo

$$y_t = \mathscr{R}_t + \gamma(1-d_t)\, Q_{\theta^-}\left(S_{t+1}, \arg\max_{a\in\mathscr{A}} Q_\theta(S_{t+1},a)\right), \tag{22}$$

donde $\gamma \in [0,1)$ es el factor de descuento y $d_t \in \{0,1\}$ un indicador que toma valor uno en la transición terminal del episodio, anulando en ese caso la contribución futura. El rasgo distintivo de la variante es que la red online identifica la acción aparentemente óptima en el estado siguiente, pero es la red objetivo la que la evalúa; este desacople reduce el sesgo de sobreestimación y evita reasignaciones excesivamente optimistas hacia combinaciones de modelos favorecidas únicamente por ruido transitorio en las estimaciones de valor.

Toda la información que utiliza el algoritmo proviene de transiciones, esto es, registros de una interacción elemental del agente con el entorno. Cada transición es una tupla $(S_t, a_t, \mathscr{R}_t, S_{t+1}, d_t)$ que almacena el estado en que se encontraba el agente, la acción ejecutada, la recompensa observada, el estado al que transitó el sistema y un indicador de término de episodio, y constituye la unidad mínima de experiencia a partir de la cual se estima la función de valor. Estas transiciones se acumulan en una memoria de repetición desde la cual, en cada paso de actualización, se muestrea aleatoriamente un mini-batch de dieciséis transiciones, lo que mejora la eficiencia muestral, al reutilizar cada experiencia en múltiples actualizaciones, y rompe la correlación temporal entre transiciones consecutivas. La red online se entrena minimizando el error cuadrático medio entre su predicción y el objetivo de Bellman,

$$\mathscr{L}(\theta) = \mathbb{E}\Big[(y_t - Q_\theta(S_t,a_t))^2\Big], \tag{23}$$

mediante el optimizador Adam, propagando el gradiente únicamente a través de $Q_\theta(S_t,a_t)$ y tratando el objetivo y_t como constante respecto de los parámetros, lo que estabiliza la dinámica de aprendizaje. La red objetivo no permanece fija, sino que se sincroniza de forma gradual mediante una actualización de Polyak,

$$\theta^- \leftarrow \tau\,\theta + (1-\tau)\,\theta^-, \tag{24}$$

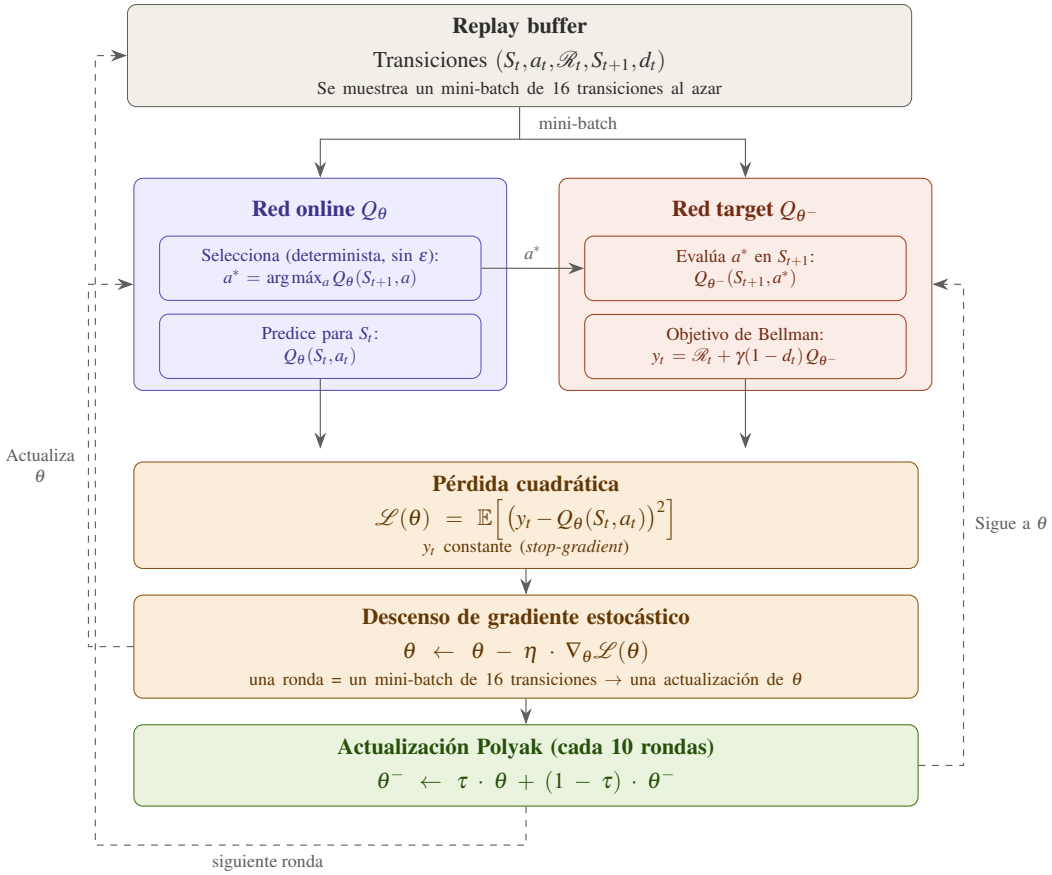


Figura 1: Paso de aprendizaje del meta-agente Double DQN sobre un mini-batch de transiciones.

aplicada cada diez pasos de gradiente; una frecuencia reducida evita el acoplamiento excesivo que tendería a colapsar la red objetivo sobre la red online. El conjunto completo de hiperparámetros del entrenamiento, junto con los esquemas de exploración y de normalización de la recompensa, se sintetiza en la Tabla 5 de la sección siguiente. La Figura 1 resume el paso de aprendizaje sobre un mini-batch que aquí se ha descrito.

3.8.6. Dinámica de entrenamiento en tres fases

La estimación de la función de valor no se realiza mediante un único bucle de optimización homogéneo, sino a través de un procedimiento estructurado en tres fases que combinan exploración aleatoria, refinamiento por descenso de gradiente y recolección periódica de experiencia bajo la política parcialmente aprendida. Esta organización responde a una restricción estructural del problema: cada combinación de subperiodo, pipeline y horizonte produce un número reducido de fechas de rebalanceo, de modo que la cantidad de transiciones disponibles

es limitada y resulta crítico aprovechar de forma estructurada cada interacción del agente con el entorno. La Figura 2 resume la secuencia de las tres fases y la recolección intercalada en la fase de refinamiento.

La primera fase, de exploración pura, construye un conjunto inicial de transiciones suficientemente diverso antes de iniciar cualquier ajuste de los parámetros de la red. Durante esta etapa el agente recorre todos los episodios disponibles y, en cada fecha de decisión, selecciona acciones de manera completamente aleatoria sobre el vocabulario, lo que equivale a fijar el parámetro de exploración en su valor máximo. Conviene precisar la unidad sobre la que opera el entrenamiento: un subperiodo es uno de los tramos anuales fuera de muestra disponibles antes del período de evaluación, y un episodio es el recorrido completo del agente sobre un subperiodo bajo un horizonte de mantención fijo, desde la primera hasta la última fecha de rebalanceo de ese tramo. El meta-agente correspondiente a un período de evaluación se entrena sobre los cuatro subperiodos anuales inmediatamente anteriores a dicho período, combinados con los tres horizontes de mantención, lo que da lugar a doce episodios de entrenamiento por agente. Las transiciones generadas se almacenan en la memoria de repetición conservando, para cada una, el estado observado, la acción ejecutada, la recompensa cruda, el estado posterior y el indicador de término de episodio. Como cada episodio aporta tantas transiciones como fechas de rebalanceo admite su horizonte, al completar los doce episodios bajo exploración aleatoria el buffer reúne setenta y dos transiciones iniciales.

Al finalizar la fase de exploración se calibra el normalizador de recompensas descrito en la sección de recompensa (Sección 3.8.4). Este normalizador opera por horizonte y aplica una transformación afín de tipo min–max que lleva las recompensas crudas a un rango aproximado entre cero y uno, utilizando únicamente las recompensas observadas durante la exploración aleatoria, momento en que el buffer contiene una muestra representativa del rango natural de valores bajo decisiones no informadas. Los valores mínimo y máximo de cada horizonte quedan fijos a partir de este punto y no se reajustan durante el resto del entrenamiento, decisión deliberada para no introducir no estacionariedad artificial en la función objetivo. Tras la calibración, las transiciones ya almacenadas se reescalan con estos parámetros, de modo que las recompensas asociadas a horizontes largos, de mayor magnitud en escala cruda, dejen de do-

minar mecánicamente el aprendizaje frente a las de horizontes más cortos.

La segunda fase, de refinamiento, constituye el núcleo del aprendizaje y ejecuta cuatrocientas rondas de actualización por descenso de gradiente, donde cada ronda corresponde al procesamiento de un único mini-batch de dieciséis transiciones extraído aleatoriamente del buffer. En cada ronda se minimiza sobre ese mini-batch la función de pérdida cuadrática entre la valoración de la red online y el objetivo de Bellman, mediante el optimizador Adam, lo que produce una única actualización de los parámetros de la red online; la red objetivo se sincroniza mediante la actualización de Polyak cada diez rondas, lo que proporciona objetivos temporales estables durante varias actualizaciones consecutivas. A diferencia de la fase de exploración, las rondas de refinamiento no generan nuevas transiciones: el agente no interactúa con el entorno en cada ronda, sino que aprende exclusivamente a partir de la experiencia ya acumulada en el buffer, la cual se repone de forma periódica mediante el mecanismo de recolección descrito a continuación.

La tercera fase, de recolección, no es una etapa separada en el tiempo, sino un mecanismo intercalado dentro del refinamiento que se activa cada cien rondas. En esos puntos, antes de continuar con el ajuste, el agente recorre nuevamente los episodios y genera transiciones adicionales utilizando la política parcialmente aprendida, con la misma regla ε-greedy pero con un valor fijo de exploración de treinta por ciento. Este nivel de exploración, suficientemente alto, evita que el buffer se concentre en un subconjunto reducido de acciones favorecidas por una política aún inmadura y preserva diversidad en la experiencia recolectada. Cada evento ejecuta un pase completo sobre los episodios y añade setenta y dos transiciones, que se incorporan al buffer y se reescalan con el normalizador ya calibrado, sin recalibrarlo. Dado que el refinamiento ejecuta cuatrocientas rondas y la recolección ocurre cada cien a partir de la centésima, se completan tres eventos que llevan el buffer de setenta y dos a doscientas ochenta y ocho transiciones. La motivación es directa: la exploración puramente aleatoria de la primera fase genera transiciones cuya distribución difiere de la que produciría la política óptima, y las re-colecciones periódicas acercan progresivamente la distribución de la experiencia disponible a la inducida por la política aprendida.

Como control de calidad del entrenamiento, al finalizar cada agente se registran diag-

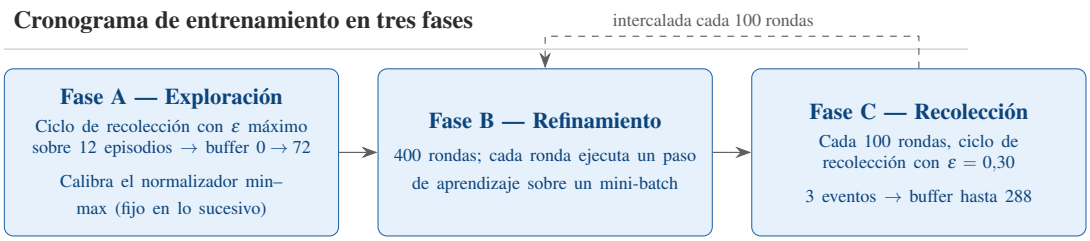


Figura 2: Dinámica de entrenamiento del meta-agente Double DQN en tres fases.

nósticos como la trayectoria de la pérdida media, cuya tendencia decreciente y estabilización evidencia convergencia adecuada; el número de acciones distintas activadas en el buffer, indicador de cobertura del vocabulario; y la concentración de las acciones más frecuentes, que permite detectar colapsos prematuros de la política hacia un único patrón de ponderación. La configuración completa del procedimiento se resume en la Tabla 5, organizada por bloques funcionales. Esta estructura en tres fases permite que el meta-agente aprenda una política contextual a partir de un número limitado de transiciones, restricción especialmente relevante en aplicaciones financieras donde la disponibilidad de fechas de decisión genuinamente fuera de muestra es estructuralmente escasa. Completado el procedimiento, el agente queda listo para su evaluación bajo el protocolo descrito en la sección siguiente.

Tabla 5: Configuración del meta-agente Double DQN.

Bloque	Parámetro	Valor
Arquitectura de la red	Dimensión del estado	37
	Capas ocultas	(64, 32)
	Dimensión de la salida	50
	Tasa de dropout	0,10
Optimización	Optimizador	Adam
	Tasa de aprendizaje η	10^{-4}
	Tamaño de mini-batch	16
	Función de pérdida	MSE frente al objetivo de Bellman
Aprendizaje secuencial	Factor de descuento γ	0,98
	Velocidad de actualización Polyak τ	0,01
	Frecuencia de actualización Polyak	cada 10 rondas
	Total de actualizaciones Polyak	40
Fase A — Exploración pura	Política de selección de acciones	aleatoria ($\varepsilon = 1{,}0$)
	Episodios totales por agente	12
	Transiciones recolectadas	72
Fase B — Refinamiento	Rondas de descenso de gradiente	400
	Modo de normalización de recompensa	min–max por horizonte
	Calibración del normalizador	al final de Fase A, fija en lo sucesivo
Fase C — Recolección	Frecuencia de recolección	cada 100 rondas
	ε fijo durante recolección	0,30
	Pases por evento de recolección	1
	Eventos totales de recolección	3
Tamaño final del buffer	Transiciones acumuladas	288

3.8.7. Protocolo fuera de muestra y construcción de señales adaptativas

La evaluación del meta-agente se desarrolla bajo un diseño temporal secuencial estricto, orientado a preservar la causalidad informacional y a medir la capacidad real de generalización del mecanismo de agregación adaptativa. Como en el resto del estudio, cada bloque temporal utiliza exclusivamente información disponible hasta la fecha correspondiente, reproduciendo las restricciones que enfrentaría un inversionista en tiempo real. Sobre la secuencia de ventanas móviles definida en la Sección 3.1, en la que cada ventana destina dos años al entrenamiento del crowd y reserva el tercero como tramo fuera de muestra, las predicciones fuera de muestra del crowd constituyen la experiencia sobre la que opera el meta-agente.

A partir de esa estructura se definen seis ejercicios de evaluación, replicados de forma idéntica en cada uno de los dos pipelines. Para cada año en que se evalúa la política adaptativa, el meta-agente se entrena con las predicciones fuera de muestra de los cuatro años inmediatamente previos, de modo que tanto la ventana de entrenamiento como el tramo de evaluación se desplazan progresivamente en el tiempo sin reutilizar información futura. Los años de evaluación final son 2020, 2021, 2022, 2023, 2024 y 2025, y la Tabla 6 resume la correspondencia con sus respectivos años de entrenamiento. Durante el entrenamiento, el agente ejecuta el ciclo de tres fases descrito en la Sección 3.8.6; durante la evaluación final no hay aprendizaje adicional, sino que la red entrenada se aplica de forma fija sobre el año no observado, sin reestimación de parámetros, sin nuevas experiencias y sin recalibración.

Tabla 6: Esquema walk-forward de las seis evaluaciones del meta-agente.

Ejercicio	Año de evaluación final	Años de experiencia para entrenamiento
1	2020	2016, 2017, 2018, 2019
2	2021	2017, 2018, 2019, 2020
3	2022	2018, 2019, 2020, 2021
4	2023	2019, 2020, 2021, 2022
5	2024	2020, 2021, 2022, 2023
6	2025	2021, 2022, 2023, 2024

La cobertura de seis años consecutivos permite analizar el comportamiento del sistema bajo regímenes macrofinancieros diferenciados, incluyendo el shock asociado a la pandemia, el episodio inflacionario posterior, el ciclo de alza de tasas y la fase de normalización subsiguiente, lo que reduce la dependencia de una única partición arbitraria de la muestra. Conviene notar

que un mismo año cumple dos funciones complementarias sin que ello genere contaminación informacional: por ejemplo, 2020 actúa como evaluación final del primer ejercicio y, a la vez, integra la experiencia de entrenamiento del segundo; cuando ingresa como experiencia lo hace estrictamente como información histórica observada, mientras que la red entrenada se evalúa siempre sobre el año posterior, no observado durante el aprendizaje.

Una vez entrenada la política, sus acciones generan en cada fecha de rebalanceo un vector de ponderaciones $w_t = (w_{1,t}, \ldots, w_{M,t})$ sobre los modelos del crowd, con $w_{m,t} \geq 0$ y $\sum_m w_{m,t} = 1$, que refleja la relevancia asignada por la política aprendida a cada modelo en esa fecha. Estas ponderaciones reemplazan la agregación simétrica de la construcción estática por una agregación dependiente del estado, y se aplican sobre la misma señal homogeneizada $x^{(h)}_{i,t,m}$ definida en la Sección 3.6, lo que preserva la coherencia dimensional entre el componente direccional y el de riesgo. El consenso adaptativo principal se define como el promedio ponderado de las señales homogeneizadas,

$$C^A_{i,t} = \sum_{m=1}^{M} w_{m,t}\, x^{(h)}_{i,t,m}, \tag{25}$$

y cuantifica la posición transversal relativa del activo respecto del crowd, ponderada según la política. El desacuerdo adaptativo principal se construye sobre la misma base como la desviación estándar ponderada de las señales individuales respecto de ese consenso,

$$D^A_{i,t} = \sqrt{\sum_{m=1}^{M} w_{m,t} \left(x^{(h)}_{i,t,m} - C^A_{i,t} \right)^2}. \tag{26}$$

Ambas magnitudes viven en la misma escala estandarizada: el consenso opera como componente direccional, capturando alineamiento relativo entre activos, y el desacuerdo como medida de riesgo informacional, capturando la dispersión interna del crowd ponderada por la política.

De esta estructura se derivan dos variantes adaptativas alternativas, paralelas a las definidas en el caso estático. La primera reemplaza las estadísticas base por medidas robustas: el consenso se calcula como la mediana ponderada de las señales homogeneizadas y el desacuerdo

como su rango intercuartílico ponderado,

$$C_{i,t}^{A,\text{med}} = \text{mediana}_w\left(x_{i,t,m}^{(h)}\right), \qquad D_{i,t}^{A,\text{iqr}} = Q_{0,75}^{w}\left(x_{i,t,m}^{(h)}\right) - Q_{0,25}^{w}\left(x_{i,t,m}^{(h)}\right), \tag{27}$$

donde el subíndice w indica que la mediana y los cuantiles se calculan sobre la distribución ponderada de pronósticos individuales bajo los pesos del agente, reduciendo la influencia de predicciones extremas en ambos componentes. La segunda variante mantiene el consenso adaptativo principal y redefine el riesgo mediante la convicción inversa,

$$D_{i,t}^{A,\text{invconv}} = \frac{D_{i,t}^{A} + \varepsilon}{\left|C_{i,t}^{A}\right| + \varepsilon}, \qquad \varepsilon > 0, \tag{28}$$

que mide la dispersión relativa del crowd por unidad de convicción direccional como una magnitud sintética adimensional.

En todos los casos, las señales adaptativas alimentan el problema de asignación descrito en la Sección 3.9 a través de un vector direccional y una matriz diagonal de riesgo informacional,

$$\mu_t = (C_{1,t}, \ldots, C_{N_t,t})', \qquad \Sigma_t = \text{diag}(D_{1,t}, \ldots, D_{N_t,t}), \tag{29}$$

donde $C_{i,t}$ y $D_{i,t}$ corresponden al consenso y al desacuerdo de la especificación considerada: la principal emplea $C_{i,t}^{A}$ y $D_{i,t}^{A}$, la robusta $C_{i,t}^{A,\text{med}}$ y $D_{i,t}^{A,\text{iqr}}$, y la de convicción $C_{i,t}^{A}$ y $D_{i,t}^{A,\text{invconv}}$. La matriz Σ_t no es una matriz de covarianzas financiera, sino un operador de penalización por dispersión relativa del crowd. La única diferencia respecto de las especificaciones estáticas es que allí las ponderaciones de los modelos son fijas, mientras que aquí los pesos w_t son seleccionados dinámicamente por la política aprendida. Al construirse ambos insumos sobre una misma representación homogeneizada, el cociente que define el problema informacional opera sobre una métrica internamente consistente, sin mezcla de unidades entre el componente direccional y el de riesgo. De este modo, la contribución del meta-agente no se evalúa sobre una métrica abstracta de predicción, sino sobre carteras efectivamente invertibles y comparables con los benchmarks del estudio.

3.9 Construcción de portafolios basados en consenso y desacuerdo

Una vez definidas las señales agregadas del crowd, el paso siguiente consiste en incorporarlas a un problema formal de asignación de portafolio. La idea central es reinterpretar esas señales como análogas a los insumos tradicionales de la teoría de selección de carteras, pero adaptadas al contexto informacional de este trabajo. El consenso por activo y fecha, expresado en escala homogeneizada, se utiliza como componente direccional del problema, mientras que el desacuerdo entre modelos, expresado en la misma escala, se interpreta como medida de riesgo informacional. Esta traducción permite aplicar reglas clásicas de optimización a un entorno en que la información no proviene de retornos históricos, sino de expectativas heterogéneas generadas por múltiples modelos predictivos, manteniendo coherencia dimensional entre los dos insumos del problema.

Sea $C_{i,t}$ la señal de consenso del activo i en la fecha t y $D_{i,t}$ su medida de desacuerdo asociada. Según la estrategia considerada, estas cantidades pueden originarse en esquemas estáticos de agregación o en versiones adaptativas obtenidas mediante aprendizaje secuencial, pero en ambos casos se construyen sobre la representación homogeneizada de las predicciones individuales. A partir de ellas se define el vector direccional informacional y la matriz diagonal de riesgo informacional,

$$\mu_t = (C_{1,t}, \ldots, C_{N_t,t})' \in \mathbb{R}^{N_t}, \qquad \Sigma_t = \operatorname{diag}(D_{1,t}, \ldots, D_{N_t,t}), \tag{30}$$

donde N_t es el número de activos disponibles en la fecha de rebalanceo. La estructura diagonal evita la estimación de covarianzas entre señales informacionales, mejora la estabilidad numérica y permite interpretar cada término como riesgo idiosincrático relativo del activo correspondiente. Es la misma matriz de riesgo informacional introducida en la subsubsección de señales adaptativas, construida a partir de medidas de desacuerdo y no de una covarianza de retornos.

Conviene precisar la naturaleza del término cuadrático antes de introducir las reglas. Las cantidades $D_{i,t}$ que ocupan la diagonal de Σ_t son desviaciones estándar del desacuerdo entre modelos, no varianzas, de modo que la forma cuadrática $\omega_t' \Sigma_t \omega_t$ se reduce a $\sum_i \omega_{i,t}^2 D_{i,t}$.

Esta expresión no es una varianza de portafolio en el sentido clásico de Markowitz, sino una agregación cuadrática de los pesos sobre los niveles de desacuerdo informacional por activo. Las denominaciones de mínima varianza informacional y media–varianza informacional empleadas a continuación se conservan por consistencia con la nomenclatura clásica, pero deben entenderse en este sentido informacional: lo que se minimiza o se penaliza es el nivel agregado de desacuerdo entre los modelos del crowd, y no una varianza de retornos.

El problema se define sobre un vector de pesos $\omega_t \in \mathbb{R}^{N_t}$, donde $\omega_{i,t}$ es la fracción de capital asignada al activo i en el rebalanceo, sujeto a restricciones de inversión larga y a una restricción de alineamiento direccional no negativo,

$$\sum_{i=1}^{N_t} \omega_{i,t} = 1, \qquad \omega_{i,t} \geq 0 \ \ \forall i, \qquad \omega_t' \mu_t \geq 0. \tag{31}$$

Las dos primeras restricciones mantienen la cartera completamente invertida y excluyen posiciones cortas. La tercera se aplica de manera homogénea a las tres reglas optimizadas y, dado que μ_t está en escala estandarizada transversal, exige que la cartera sobrepondere en promedio los activos cuyo consenso se ubica al menos en el nivel medio de la sección cruzada, prohibiendo asignaciones cuyo alineamiento direccional agregado con el crowd sea negativo. Esta condición es inactiva cuando el consenso presenta dispersión transversal con activos por encima de la media, y solo se vuelve vinculante en regímenes de fuerte concentración direccional negativa, en los que la cartera converge a soluciones de esquina sobre los activos relativamente menos desfavorables.

Sobre este marco se implementan tres reglas de asignación. La primera es el portafolio de mínima varianza informacional, que minimiza la discrepancia informacional agregada,

$$\min_{\omega_t} \ \ \omega_t' \Sigma_t \omega_t, \tag{32}$$

sujeto a las restricciones comunes, seleccionando la combinación de activos con menor desacuerdo agregado entre las que satisfacen un alineamiento direccional al menos neutral. La

segunda es el portafolio de máximo ratio de Sharpe informacional,

$$\max_{\omega_t} \ \frac{\omega_t' \mu_t}{\sqrt{\omega_t' \Sigma_t \omega_t}}, \tag{33}$$

que escoge la combinación con mayor alineamiento direccional con el consenso por unidad de discrepancia informacional. Como μ_t está en escala estandarizada, el numerador vive en el mismo espacio que el denominador, lo que preserva la coherencia dimensional del cociente. La tasa libre de riesgo no aparece en este numerador, pues restar una cantidad expresada en escala de retorno a un vector expresado en escala estandarizada mezclaría unidades sin interpretación económica directa; dicha tasa se emplea, en cambio, en el ratio de Sharpe realizado utilizado como métrica de desempeño fuera de muestra. La tercera regla es una versión media–varianza con aversión al riesgo explícita,

$$\max_{\omega_t} \ \omega_t' \mu_t - \lambda \, \omega_t' \Sigma_t \omega_t, \qquad \lambda > 0, \tag{34}$$

donde λ regula el peso del componente de desacuerdo frente al direccional: valores mayores inducen carteras más concentradas en activos de bajo desacuerdo, y valores menores privilegian el alineamiento con el consenso. La implementación empírica utiliza $\lambda = 3$, valor elegido para equilibrar ambos componentes en el espacio común estandarizado.

En la implementación numérica, los tres problemas se resuelven mediante el algoritmo de programación cuadrática secuencial sobre múltiples puntos de partida aleatorios en el simplex, conservando la solución factible con mejor valor objetivo. En el caso degenerado en que ningún punto de partida converge a una solución que satisfaga simultáneamente las restricciones de presupuesto, no negatividad y alineamiento direccional no negativo, la implementación retorna la asignación equiponderada sobre los activos disponibles. Este mecanismo de respaldo opera como solución conservadora frente a regímenes en que el conjunto factible bajo la restricción $\omega_t' \mu_t \geq 0$ colapsa numéricamente. En cada fecha de decisión los pesos se determinan utilizando solo información disponible en t; resuelto el problema, la cartera se implementa desde la siguiente fecha observable y permanece constante hasta el rebalanceo correspondiente al horizonte, de modo que la señal de inversión antecede siempre a los retornos usados en la

evaluación. Las tres reglas son agnósticas respecto del origen de las señales $C_{i,t}$ y $D_{i,t}$: la diferencia entre estrategias proviene exclusivamente de cómo se construyen estos insumos, y la correspondencia precisa entre cada par de insumos y la estrategia que induce se detalla en la Sección 3.11.

3.10 Portafolios de referencia

Para evaluar el desempeño de las estrategias informacionales se consideran portafolios de referencia que representan reglas ampliamente utilizadas en la selección de carteras. La comparación con ellos permite determinar si el contenido informacional del crowd y la capa adaptativa generan valor económico adicional respecto de estrategias construidas con retornos históricos o de exposición pasiva al mercado. Los benchmarks históricos se estiman con una memoria retrospectiva de la misma longitud que el horizonte de inversión evaluado: en cada fecha de rebalanceo, el vector de retornos esperados históricos $\mu_t^{(r)}$ y la matriz de covarianzas $\Sigma_t^{(r)}$ se estiman con los últimos h días hábiles previos a la decisión. Esta elección privilegia el alineamiento entre el horizonte de predicción y la ventana histórica del benchmark; si bien puede introducir mayor error de estimación en horizontes cortos, especialmente en la matriz de covarianzas, mantiene una referencia estrictamente alineada con la escala temporal de cada ejercicio.

El primer benchmark es el portafolio histórico de mínima varianza, que minimiza $\omega_t' \Sigma_t^{(r)} \omega_t$ sujeto a las restricciones de presupuesto, no negatividad y retorno esperado histórico no negativo. El segundo es el portafolio histórico de máximo ratio de Sharpe clásico,

$$\max_{\omega_t} \ \frac{\omega_t' \mu_t^{(r)} - r_{f,h}}{\sqrt{\omega_t' \Sigma_t^{(r)} \omega_t}}, \tag{35}$$

donde, a diferencia del Sharpe informacional, los insumos están en escala de retorno y la tasa libre de riesgo $r_{f,h}$, transformada al horizonte h, sí ingresa en el numerador. Ambos benchmarks optimizados emplean las mismas restricciones que los portafolios informacionales, excluyendo soluciones con retorno esperado histórico agregado negativo, lo que permite una comparación homogénea entre metodologías. El tercer benchmark es el portafolio de pesos iguales,

$\omega_{i,t} = 1/N_t$, que no requiere estimación de parámetros y constituye una referencia parsimoniosa y robusta frente al error de estimación. Finalmente, se incorpora el índice S&P 500 como benchmark de mercado, interpretado como una estrategia pasiva de inversión amplia en renta variable estadounidense.

Todos los benchmarks se rebalancean en las mismas fechas que las estrategias informacionales para el horizonte correspondiente, y sus insumos históricos se recalculan en cada rebalanceo con información disponible hasta la fecha de decisión. Las métricas reportadas corresponden a desempeño bruto antes de costos explícitos de transacción, por lo que los resultados deben interpretarse como comparaciones relativas entre reglas de asignación bajo un mismo conjunto de supuestos operacionales. En conjunto, estos benchmarks permiten evaluar si el desempeño de las estrategias propuestas proviene efectivamente del contenido informacional del crowd y de la adaptación dinámica entre modelos, o si resultados similares podrían alcanzarse mediante reglas históricas estándar o exposición pasiva al mercado.

3.11 Portafolios evaluados

Las subsecciones anteriores definen por separado los problemas de optimización, los benchmarks de referencia y las distintas formas de construir las señales de consenso y desacuerdo, tanto estáticas como adaptativas. Esta subsección consolida esas definiciones e identifica los portafolios efectivamente evaluados en el estudio empírico. Cada portafolio informacional se obtiene combinando un par específico de señales $(C_{i,t}, D_{i,t})$ con una de las tres reglas de optimización. La especificación adaptativa principal se evalúa bajo los tres objetivos, por tratarse de la línea central del estudio, mientras que las variantes adaptativas de robustez y las tres especificaciones estáticas se evalúan exclusivamente bajo el objetivo de máximo Sharpe informacional, que resume de manera más directa la disyuntiva entre alineamiento direccional con el consenso y discrepancia informacional agregada. Restringir las variantes secundarias a un único objetivo responde a un criterio de parsimonia: permite aislar el efecto del cambio en la regla de agregación sobre una métrica común, sin multiplicar el número de configuraciones reportadas.

A estos portafolios informacionales se suman los benchmarks históricos optimizados y

los benchmarks no optimizados, conformando un total de doce estrategias. La Tabla 7 resume el inventario completo, indicando para cada estrategia el origen de los insumos y la regla de optimización aplicada.

Tabla 7: Inventario de las doce estrategias evaluadas.

Familia	$C_{i,t}$	$D_{i,t}$	Optimización
Benchmarks			
S&P 500	—	—	Exposición pasiva
Pesos iguales	—	—	$\omega_{i,t} = 1/N_t$
Histórico mínima varianza	Media histórica	Covarianza histórica	Mínima varianza
Histórico máximo Sharpe	Media histórica	Covarianza histórica	Máximo Sharpe
Estáticos			
Estática principal	$C_{i,t}^{\text{mean}}$	$D_{i,t}^{\text{std}}$	Máximo Sharpe
Estática robusta	$C_{i,t}^{\text{med}}$	$D_{i,t}^{\text{iqr}}$	Máximo Sharpe
Estática convicción	$C_{i,t}^{\text{mean}}$	$D_{i,t}^{\text{invconv}}$	Máximo Sharpe
Adaptativos			
Adaptativa principal	$C_{i,t}^{A}$	$D_{i,t}^{A}$	Máximo Sharpe
Adaptativa principal	$C_{i,t}^{A}$	$D_{i,t}^{A}$	Mínima varianza
Adaptativa principal	$C_{i,t}^{A}$	$D_{i,t}^{A}$	Media–varianza
Adaptativa robusta	$C_{i,t}^{A,\text{med}}$	$D_{i,t}^{A,\text{iqr}}$	Máximo Sharpe
Adaptativa convicción	$C_{i,t}^{A}$	$D_{i,t}^{A,\text{invconv}}$	Máximo Sharpe

En todos los ejercicios se mantienen constantes las restricciones de inversión, la frecuencia de rebalanceo, los horizontes de mantención y el protocolo fuera de muestra. Por lo tanto, cualquier diferencia de desempeño entre las estrategias informacionales puede atribuirse exclusivamente a la forma en que las predicciones del crowd se agregan en señales utilizables para la asignación de activos, mientras que la comparación frente a los benchmarks permite evaluar si ese contenido informacional genera valor económico adicional respecto de reglas históricas estándar o de exposición pasiva al mercado.

4 Resultados y discusión

4.1 Calidad predictiva del crowd

Antes de evaluar las carteras conviene caracterizar la materia prima del sistema: la calidad de los pronósticos individuales que conforman el crowd. La Tabla 8 reporta, para cada uno de los nueve modelos, la raíz del error cuadrático medio (RMSE) y la precisión de signo (ACC) definidas en la Sección 3.5, promediadas sobre los diez períodos fuera de muestra y desagregadas por horizonte y pipeline.

Tabla 8: Calidad predictiva del crowd: raíz del error cuadrático medio (RMSE) y precisión de signo (ACC) por modelo, horizonte y pipeline.

	$h = 20$		$h = 60$		$h = 120$	
Modelo	RMSE	ACC	RMSE	ACC	RMSE	ACC
Panel A — Pipeline RAW						
MLP	0,0898	0,535	0,1419	0,565	0,1851	0,640
LSTM	0,0809	0,528	0,1361	0,580	0,1840	0,635
GARCH-LSTM	0,0878	0,513	0,1402	0,497	0,1969	0,508
EGARCH-LSTM	0,0954	0,515	0,1394	0,523	0,2035	0,526
RF	0,0797	0,551	0,1342	0,573	0,1879	0,645
XGBoost	0,0831	0,532	0,1350	0,591	0,1739	0,682
HGB	0,0814	0,559	0,1327	0,599	0,1742	0,667
MA	0,1092	0,509	0,1813	0,538	0,2469	0,552
EMA	0,1026	0,515	0,1791	0,526	0,2500	0,552
Panel B — Pipeline WW						
MLP	0,0879	0,519	0,1527	0,550	0,2109	0,602
LSTM	0,0845	0,526	0,1461	0,577	0,2041	0,623
GARCH-LSTM	0,0815	0,528	0,1375	0,562	0,1933	0,591
EGARCH-LSTM	0,0791	0,513	0,1362	0,552	0,1886	0,588
RF	0,0852	0,534	0,1456	0,563	0,2039	0,604
XGBoost	0,0825	0,525	0,1433	0,553	0,2018	0,576
HGB	0,0808	0,544	0,1420	0,574	0,2005	0,598
MA	0,1063	0,509	0,1777	0,535	0,2427	0,552
EMA	0,1007	0,517	0,1763	0,526	0,2504	0,557

Nota: Cada valor es el promedio sobre los diez períodos de evaluación fuera de muestra.

Tres regularidades emergen del panel correspondiente al pipeline RAW. En primer lugar, la RMSE crece de forma monótona con el horizonte en todos los modelos, lo que es esperable porque el retorno acumulado a 120 días posee una varianza mayor que el de 20 días. En segundo lugar, dentro de cada horizonte los modelos de árboles —RF, XGBoost y HGB— y la LSTM exhiben los menores errores, mientras que las dos reglas de promedio móvil, MA y EMA, registran sistemáticamente los mayores, coherente con su carácter de referencias parsimoniosas. En tercer lugar, los híbridos de volatilidad GARCH-LSTM y EGARCH-LSTM ocupan una posición intermedia en magnitud de error.

El patrón de la precisión de signo es más informativo para los fines de este trabajo. En términos generales, la mayoría de las combinaciones modelo–horizonte se ubica por encima del umbral de azar de 0,5, y la capacidad direccional aumenta de manera marcada hacia el horizonte de 120 días. En este horizonte, los modelos de árboles alcanzan los valores más altos, con XGBoost en 0,682 y HGB en 0,667, seguidos por RF (0,645), el perceptrón multicapa (MLP, 0,640) y la LSTM (0,635). Esta superioridad de los métodos no lineales en la anticipación del signo es consistente con la evidencia de que árboles y redes neuronales capturan en los retornos

interacciones que las especificaciones lineales no recogen (Gu et al., 2020). Los híbridos de volatilidad y las reglas de promedio móvil, en cambio, se mantienen cerca del azar en dirección, entre 0,51 y 0,55, lo que sugiere que su aporte al crowd es menos direccional y se vincula más a la caracterización del régimen de riesgo. Esta heterogeneidad —modelos fuertes en dirección junto a modelos débiles en dirección pero útiles por otras razones— es precisamente la diversidad que justifica combinar el conjunto en lugar de seleccionar un único ganador, en línea con el argumento de diversificación de la combinación de pronósticos (Timmermann, 2006).

La comparación entre pipelines en la Tabla 8 anticipa un resultado central de las secciones siguientes. El tratamiento WW —filtrado wavelet y winsorización del objetivo— no mejora de manera uniforme la calidad predictiva, sino que produce un efecto heterogéneo según la familia de modelos. Por un lado, deteriora la precisión de signo de los modelos de árboles y de las redes simples: la ACC de XGBoost a 120 días cae de 0,682 en RAW a 0,576 en WW, y la de HGB de 0,667 a 0,598. Por otro lado, beneficia a los híbridos de volatilidad: la RMSE del EGARCH-LSTM a 120 días disminuye de 0,2035 a 0,1886 y su ACC sube de 0,526 a 0,588. Una interpretación plausible es que el suavizado del objetivo reduce componentes de la serie que los modelos de árboles utilizaban para ordenar direccionalmente los activos, mientras que puede favorecer a modelos diseñados para capturar dinámicas más persistentes de varianza. Dado que la capacidad direccional del crowd se concentra especialmente en los modelos de árboles, este deterioro selectivo anticipa una posible explicación del menor desempeño económico observado posteriormente en el pipeline WW. En este sentido, los resultados son consistentes con la evidencia de que el preprocesamiento y la reducción de ruido pueden alterar de forma material —y no siempre favorable— el desempeño de los sistemas predictivos en finanzas (Chan Phooi M'ng & Mehralizadeh, 2016; Nobre & Neves, 2019).

4.2 Convergencia y diagnósticos del meta-agente

La capa adaptativa del marco consta de doce meta-agentes Double DQN, uno por cada combinación de los seis años de evaluación y los dos pipelines, entrenados bajo el procedimiento de tres fases de la Sección 3.8.6. Antes de analizar las carteras que inducen, es necesario verificar que el aprendizaje fue numéricamente sano, esto es, que la pérdida se redujo de forma

estable y que la política exploró el vocabulario de acciones sin colapsar prematuramente. La Tabla 9 resume los diagnósticos de entrenamiento de los doce agentes.

Tabla 9: Diagnósticos de entrenamiento de los doce meta-agentes Double DQN.

Año	Pipeline	Pérdida inicial	Pérdida final	Pérdida mínima	Acciones distintas	Conc. top-10 (%)
2020	RAW	1,360	0,599	0,160	49/50	58,3
	WW	1,320	0,533	0,133	49/50	56,6
2021	RAW	1,960	0,876	0,253	49/50	60,4
	WW	1,931	0,741	0,143	49/50	55,9
2022	RAW	1,326	0,636	0,257	49/50	59,0
	WW	1,209	0,494	0,197	49/50	57,6
2023	RAW	1,556	0,677	0,201	49/50	58,3
	WW	1,280	0,554	0,186	49/50	58,0
2024	RAW	1,465	0,620	0,188	49/50	60,1
	WW	1,160	0,495	0,221	49/50	58,7
2025	RAW	1,419	0,833	0,156	49/50	59,4
	WW	1,247	0,464	0,162	49/50	58,7

Nota: La pérdida inicial y final corresponden a la función de pérdida en la primera y la última ronda de refinamiento, y la pérdida mínima al menor valor alcanzado durante esa fase; las acciones distintas y la concentración de las diez acciones más frecuentes se calculan sobre la memoria de repetición final.

La trayectoria de la pérdida es consistente con un entrenamiento numéricamente estable en todos los casos. La pérdida media inicial, calculada sobre las primeras rondas de refinamiento, se sitúa entre 1,16 y 1,96 según el combo, y desciende de forma consistente hasta valores finales de entre 0,46 y 0,88, con mínimos a lo largo del entrenamiento de entre 0,13 y 0,26. Esta reducción sostenida y su posterior estabilización son la firma esperada de un entrenamiento estable, atribuible a los mecanismos que el Deep Q-Network introduce con ese fin —la memoria de repetición y la red objetivo separada (Mnih et al., 2015)— complementados por el desacople entre selección y evaluación de la acción propio de la variante Double DQN, que evita la sobreestimación sistemática de los valores (van Hasselt et al., 2016). En todos los combos la memoria de repetición alcanzó las 288 transiciones previstas tras las tres re-colecciones, y las recompensas finales resultaron finitas y sin valores degenerados.

Los diagnósticos de cobertura del espacio de acciones descartan un colapso prematuro de la política. En los doce agentes el buffer terminó conteniendo 49 de las 50 acciones del vocabulario, lo que indica que la exploración recorrió casi por completo el repertorio de combinaciones de modelos. Al mismo tiempo, la concentración de las diez acciones más frecuentes se mantuvo en un rango acotado de entre 55,9 % y 60,4 %, un nivel que refleja una política selectiva —que efectivamente prefiere ciertas combinaciones— pero no degenerada hacia un

único patrón. De hecho, en once de los doce combos la acción más visitada fue la misma, lo que sugiere que la política aprende a concentrarse en una región recurrente del espacio de ponderaciones, con independencia del régimen y del pipeline, en lugar de saltar erráticamente entre configuraciones. Esta estabilidad es deseable, pues indica que el agente identifica un núcleo consistente de modelos informativos y reasigna su atención de forma controlada alrededor de él.

4.3 Desempeño económico por pipeline y horizonte

Verificada la convergencia, se evalúa el desempeño económico de las doce estrategias definidas en la Sección 3.11. Las Tablas 10 y 11 reportan, para los pipelines RAW y WW respectivamente, la equivalencia geométrica y los promedios de volatilidad, ratio de Sharpe, ratio de Sortino y ratio de Calmar, organizados en cuatro bloques: un bloque agregado y los tres horizontes de mantención. Conviene precisar la diferencia entre las dos formas de resumir el retorno: mientras los ratios se promedian sobre los seis períodos a partir de retornos logarítmicos, la equivalencia geométrica resume la secuencia de retornos por período en una tasa media compuesta. Para ello, el retorno logarítmico de cada período se traduce a su retorno aritmético equivalente y luego se calcula el promedio geométrico, evitando que años extremos dominen la interpretación como ocurriría con un promedio aritmético simple.

En el pipeline RAW (Tabla 10) las estrategias informacionales —tanto estáticas como adaptativas— superan a los benchmarks clásicos. En términos agregados, el ratio de Sharpe del índice S&P 500 es 0,791 y el de la cartera equiponderada 0,531, mientras que los benchmarks históricos optimizados quedan claramente por detrás, con 0,431 la mínima varianza y 0,439 el máximo Sharpe histórico, un reflejo de la conocida fragilidad de los insumos de media–varianza estimados sobre ventanas cortas. Frente a ellos, las tres estrategias estáticas alcanzan ratios de Sharpe agregados de entre 0,951 y 0,983, y entre las adaptativas la principal de máximo Sharpe obtiene 0,891, la de convicción 0,862 y la de media–varianza 0,796. Con ello, las tres estáticas y las adaptativas de máximo Sharpe (0,891) y de convicción (0,862) superan con claridad a la exposición pasiva al mercado (0,791); la adaptativa de media–varianza (0,796) la iguala en la práctica, mientras que la adaptativa robusta (0,785) y, sobre todo, la de mínima varianza (0,608)

Tabla 10: Desempeño de las doce estrategias en el pipeline RAW.

Horizonte	Estrategia	Equiv. geom.	Vol.	Sharpe	Sortino	Calmar
Agregado	S&P 500	0,135	0,191	0,791	1,083	1,820
	Pesos iguales	0,098	0,179	0,531	0,757	1,234
	Hist. mínima varianza	0,097	0,155	0,431	0,590	1,070
	Hist. máximo Sharpe	0,104	0,212	0,439	0,669	1,274
	Estática principal	0,223	0,211	0,951	1,484	2,017
	Estática robusta	0,227	0,216	0,983	1,540	2,100
	Estática convicción	0,264	0,240	0,957	1,472	2,117
	Adaptativa principal (máx. Sharpe)	0,220	0,223	0,891	1,382	1,863
	Adaptativa principal (mín. var.)	0,109	0,172	0,608	0,864	1,451
	Adaptativa principal (media–var.)	0,256	0,277	0,796	1,195	1,699
	Adaptativa robusta	0,213	0,237	0,785	1,183	1,756
	Adaptativa convicción	0,252	0,254	0,862	1,313	1,853
$h = 20$	S&P 500	0,135	0,191	0,791	1,083	1,820
	Pesos iguales	0,098	0,179	0,531	0,757	1,234
	Hist. mínima varianza	0,112	0,152	0,445	0,630	1,062
	Hist. máximo Sharpe	0,087	0,211	0,273	0,374	0,714
	Estática principal	0,189	0,211	0,836	1,337	1,980
	Estática robusta	0,198	0,217	0,900	1,440	2,107
	Estática convicción	0,201	0,244	0,749	1,146	1,957
	Adaptativa principal (máx. Sharpe)	0,196	0,215	0,854	1,314	1,859
	Adaptativa principal (mín. var.)	0,115	0,173	0,655	0,925	1,610
	Adaptativa principal (media–var.)	0,187	0,265	0,647	0,968	1,414
	Adaptativa robusta	0,203	0,233	0,808	1,272	2,243
	Adaptativa convicción	0,185	0,244	0,693	1,044	1,546
$h = 60$	S&P 500	0,135	0,191	0,791	1,083	1,820
	Pesos iguales	0,098	0,179	0,531	0,757	1,234
	Hist. mínima varianza	0,106	0,153	0,425	0,549	1,047
	Hist. máximo Sharpe	0,100	0,201	0,535	0,894	2,081
	Estática principal	0,239	0,221	1,015	1,549	1,970
	Estática robusta	0,216	0,234	0,900	1,348	1,706
	Estática convicción	0,265	0,252	1,012	1,588	2,089
	Adaptativa principal (máx. Sharpe)	0,210	0,243	0,782	1,183	1,536
	Adaptativa principal (mín. var.)	0,105	0,173	0,560	0,800	1,240
	Adaptativa principal (media–var.)	0,223	0,316	0,610	0,865	1,288
	Adaptativa robusta	0,198	0,258	0,670	0,938	1,341
	Adaptativa convicción	0,235	0,279	0,736	1,079	1,486
$h = 120$	S&P 500	0,135	0,191	0,791	1,083	1,820
	Pesos iguales	0,098	0,179	0,531	0,757	1,234
	Hist. mínima varianza	0,065	0,159	0,421	0,592	1,102
	Hist. máximo Sharpe	0,116	0,223	0,509	0,739	1,028
	Estática principal	0,233	0,202	1,002	1,565	2,102
	Estática robusta	0,258	0,196	1,149	1,833	2,486
	Estática convicción	0,303	0,224	1,109	1,683	2,306
	Adaptativa principal (máx. Sharpe)	0,248	0,210	1,038	1,647	2,193
	Adaptativa principal (mín. var.)	0,107	0,170	0,609	0,867	1,503
	Adaptativa principal (media–var.)	0,339	0,252	1,129	1,752	2,396
	Adaptativa robusta	0,225	0,221	0,876	1,339	1,684
	Adaptativa convicción	0,324	0,237	1,157	1,816	2,525

Tabla 11: Desempeño de las doce estrategias en el pipeline WW.

Horizonte	Estrategia	Equiv. geom.	Vol.	Sharpe	Sortino	Calmar
Agregado	S&P 500	0,135	0,191	0,791	1,083	1,820
	Pesos iguales	0,098	0,179	0,531	0,757	1,234
	Hist. mínima varianza	0,097	0,155	0,431	0,590	1,070
	Hist. máximo Sharpe	0,104	0,212	0,439	0,669	1,274
	Estática principal	0,111	0,210	0,517	0,730	1,297
	Estática robusta	0,111	0,218	0,499	0,675	1,268
	Estática convicción	0,133	0,236	0,500	0,727	1,270
	Adaptativa principal (máx. Sharpe)	0,077	0,219	0,318	0,421	0,880
	Adaptativa principal (mín. var.)	0,088	0,175	0,508	0,716	1,265
	Adaptativa principal (media–var.)	0,096	0,263	0,257	0,352	0,754
	Adaptativa robusta	0,114	0,253	0,314	0,425	0,754
	Adaptativa convicción	0,082	0,244	0,255	0,335	0,777
$h = 20$	S&P 500	0,135	0,191	0,791	1,083	1,820
	Pesos iguales	0,098	0,179	0,531	0,757	1,234
	Hist. mínima varianza	0,112	0,152	0,445	0,630	1,062
	Hist. máximo Sharpe	0,087	0,211	0,273	0,374	0,714
	Estática principal	0,116	0,215	0,558	0,813	1,365
	Estática robusta	0,124	0,220	0,563	0,795	1,418
	Estática convicción	0,151	0,243	0,591	0,868	1,219
	Adaptativa principal (máx. Sharpe)	0,095	0,223	0,403	0,546	0,975
	Adaptativa principal (mín. var.)	0,106	0,176	0,597	0,840	1,356
	Adaptativa principal (media–var.)	0,144	0,265	0,430	0,600	0,942
	Adaptativa robusta	0,102	0,236	0,334	0,424	0,744
	Adaptativa convicción	0,112	0,251	0,372	0,503	0,959
$h = 60$	S&P 500	0,135	0,191	0,791	1,083	1,820
	Pesos iguales	0,098	0,179	0,531	0,757	1,234
	Hist. mínima varianza	0,106	0,153	0,425	0,549	1,047
	Hist. máximo Sharpe	0,100	0,201	0,535	0,894	2,081
	Estática principal	0,110	0,214	0,480	0,669	1,270
	Estática robusta	0,099	0,221	0,415	0,553	1,118
	Estática convicción	0,135	0,239	0,487	0,730	1,478
	Adaptativa principal (máx. Sharpe)	0,049	0,217	0,202	0,267	0,816
	Adaptativa principal (mín. var.)	0,073	0,176	0,422	0,600	1,095
	Adaptativa principal (media–var.)	0,063	0,247	0,160	0,229	0,741
	Adaptativa robusta	0,135	0,237	0,447	0,620	0,987
	Adaptativa convicción	0,038	0,240	0,096	0,128	0,697
$h = 120$	S&P 500	0,135	0,191	0,791	1,083	1,820
	Pesos iguales	0,098	0,179	0,531	0,757	1,234
	Hist. mínima varianza	0,065	0,159	0,421	0,592	1,102
	Hist. máximo Sharpe	0,116	0,223	0,509	0,739	1,028
	Estática principal	0,105	0,201	0,513	0,708	1,257
	Estática robusta	0,110	0,212	0,518	0,679	1,268
	Estática convicción	0,108	0,225	0,422	0,582	1,113
	Adaptativa principal (máx. Sharpe)	0,083	0,216	0,350	0,451	0,849
	Adaptativa principal (mín. var.)	0,086	0,174	0,505	0,708	1,343
	Adaptativa principal (media–var.)	0,074	0,277	0,180	0,227	0,580
	Adaptativa robusta	0,094	0,286	0,161	0,231	0,532
	Adaptativa convicción	0,089	0,242	0,299	0,373	0,675

quedan por debajo.

Un rasgo que merece subrayarse es que, en el agregado, las estrategias estáticas no solo son competitivas sino que exhiben los ratios de Sharpe más altos, por encima de la adaptativa principal. Lejos de constituir un resultado adverso, esto es consistente con la robustez empírica que la literatura de combinación de pronósticos atribuye a las reglas de agregación simples, notoriamente difíciles de superar mediante esquemas que estiman ponderaciones (Timmermann, 2006). La agregación estática opera, por tanto, como un rival exigente y no como un punto de comparación débil, lo que vuelve más significativo el hecho de que la capa adaptativa lo iguale o supere en horizontes y regímenes específicos, como se documenta a continuación.

La descomposición por horizonte revela que el valor de la agregación adaptativa depende del horizonte de inversión. En el horizonte de 20 días la estrategia estática en su versión robusta lidera con un ratio de Sharpe de 0,900, estando en segundo lugar la adaptativa principal de máximo Sharpe (0,854). En el horizonte de 60 días el liderazgo sigue en las estáticas, con la estática principal y la de convicción en torno a 1,01. En el horizonte de 120 días, en cambio, la capa adaptativa alcanza su mejor desempeño relativo: la adaptativa de convicción registra el ratio de Sharpe promedio más alto del estudio (1,157), seguida por la adaptativa de media–varianza (1,129) y la adaptativa principal de máximo Sharpe (1,038), todas por encima del mercado (0,791) y a la altura de las mejores estáticas. La equivalencia geométrica a 120 días confirma el patrón: la adaptativa de media–varianza (0,339) y la de convicción (0,324) exhiben la mayor tasa de crecimiento compuesto del capital. Que la ventaja de la adaptación se concentre en el horizonte largo es coherente con la evidencia de que el efecto del desacuerdo sobre los retornos es más marcado en horizontes de uno a tres años (Yu, 2011), lo que refuerza la pertinencia económica del mecanismo en esa escala temporal.

El pipeline WW (Tabla 11) ofrece un contraste nítido: el tratamiento del objetivo deteriora el desempeño de forma generalizada y golpea con especial dureza a las estrategias adaptativas. En el agregado, la adaptativa principal de máximo Sharpe cae a un ratio de Sharpe de 0,318 y la de convicción a 0,255, ambas por debajo no solo del mercado sino incluso de la cartera equiponderada (0,531); las estáticas también retroceden, a un rango de entre 0,499 y 0,517. El deterioro es más severo en el horizonte largo: a 120 días la adaptativa de convicción,

que en RAW lideraba con 1,157, desciende a 0,299. La estrategia informacional que menos se deteriora bajo WW es la adaptativa de mínima varianza, cuyo ratio de Sharpe agregado desciende de 0,608 a 0,508, frente a la caída de 0,951 a 0,517 de la estática principal, lo que es esperable porque su objetivo descansa en el nivel de desacuerdo y no en la señal direccional que el suavizado degrada. Este resultado conecta directamente con la calidad predictiva documentada en la Sección 4.1: al erosionar la precisión de signo de los modelos de árboles —el motor direccional del crowd—, el filtrado del objetivo priva a las estrategias de máximo Sharpe y de convicción de su insumo más valioso. La evidencia es así coherente con la idea de que la reducción de ruido en finanzas puede resultar contraproducente cuando elimina señal genuina junto con el ruido (Chan Phooi M'ng & Mehralizadeh, 2016; Nobre & Neves, 2019).

En conjunto, la comparación entre pipelines es inequívoca: entre las estrategias informacionales, el pipeline RAW domina al WW de manera uniforme a lo largo de horizontes y especificaciones. Por esta razón, el análisis detallado de la subsección siguiente se concentra en el pipeline RAW. Dentro de él, se adopta el horizonte de 120 días como caso focal por una doble justificación. Primero, es el horizonte en que la capa adaptativa alcanza su mayor aporte económico dentro del diseño empírico. Segundo, aunque este horizonte es más corto que los tramos anuales usuales en parte de la literatura sobre desacuerdo, corresponde a la escala más larga considerada en este estudio y es precisamente donde la señal informacional alcanza mayor contenido económico (Yu, 2011). Como las Tablas 10 y 11 reportan ya todos los horizontes, esta elección no oculta información, sino que permite profundizar en la configuración donde el aporte económico de las estrategias informacionales es más visible. La descomposición por régimen que sigue debe interpretarse, por tanto, como un análisis focal de la configuración más informativa y no como una selección aislada de resultados.

4.4 Profundización en la configuración de mejor comportamiento: pipeline RAW, horizonte de 120 días

Esta subsección examina con mayor detalle el comportamiento de las estrategias en la configuración de mejor desempeño —pipeline RAW y horizonte de 120 días—, descomponiendo los resultados por año de evaluación para distinguir el aporte de cada estrategia según

el régimen de mercado. La Tabla 12 reporta el ratio de Sharpe por año y la Tabla 13 el retorno aritmético correspondiente. La tabla de ratios cierra con el promedio sobre los seis años; la de retornos, con la equivalencia geométrica, que resume la serie de forma más informativa que un promedio aritmético al incorporar el efecto de la capitalización. Estos años cubren regímenes contrastantes: el shock de la pandemia en 2020, la recuperación de 2021, el mercado bajista e inflacionario de 2022 y la fase de normalización de 2023 a 2025.

Tabla 12: Ratio de Sharpe por año en la configuración ganadora.

Estrategia	2020	2021	2022	2023	2024	2025	Promedio
S&P 500	0,314	1,685	−0,884	1,348	1,702	0,582	0,791
Pesos iguales	0,217	1,560	−0,465	0,556	0,763	0,552	0,531
Hist. mínima varianza	−0,089	1,287	−0,419	−0,581	1,623	0,708	0,421
Hist. máximo Sharpe	0,187	1,430	−0,278	0,187	1,311	0,219	0,509
Estática principal	0,472	1,642	0,600	1,260	1,347	0,689	1,002
Estática robusta	0,487	1,924	0,628	1,507	1,648	0,699	1,149
Estática convicción	0,595	1,824	1,046	1,403	0,867	0,920	1,109
Adaptativa principal (máx. Sharpe)	0,603	1,867	0,118	1,628	1,345	0,665	1,038
Adaptativa principal (mín. var.)	0,264	1,655	−0,429	0,657	0,986	0,524	0,609
Adaptativa principal (media–var.)	0,867	2,248	0,208	1,843	0,968	0,642	1,129
Adaptativa robusta	0,318	1,651	−0,097	1,501	1,161	0,723	0,876
Adaptativa convicción	0,836	2,159	0,208	1,807	1,208	0,725	1,157

Tabla 13: Retorno aritmético por año en la configuración ganadora.

Estrategia	2020	2021	2022	2023	2024	2025	Equiv. geom.
S&P 500	0,125	0,252	−0,167	0,234	0,276	0,159	0,135
Pesos iguales	0,090	0,213	−0,070	0,112	0,127	0,135	0,098
Hist. mínima varianza	−0,023	0,166	−0,043	−0,025	0,212	0,135	0,065
Hist. máximo Sharpe	0,083	0,280	−0,045	0,082	0,252	0,079	0,116
Estática principal	0,183	0,372	0,178	0,260	0,245	0,172	0,233
Estática robusta	0,179	0,410	0,187	0,344	0,295	0,156	0,258
Estática convicción	0,232	0,497	0,339	0,310	0,208	0,255	0,303
Adaptativa principal (máx. Sharpe)	0,234	0,383	0,059	0,427	0,260	0,161	0,248
Adaptativa principal (mín. var.)	0,103	0,226	−0,057	0,123	0,148	0,121	0,107
Adaptativa principal (media–var.)	0,406	0,542	0,097	0,663	0,244	0,174	0,339
Adaptativa robusta	0,115	0,409	0,002	0,443	0,277	0,164	0,225
Adaptativa convicción	0,367	0,482	0,091	0,613	0,262	0,196	0,324

La descomposición por régimen revela diferencias entre las estrategias que el agregado tiende a enmascarar. En el año del shock pandémico (2020), las estrategias adaptativas ofrecen la mejor combinación de protección y retorno: la adaptativa de media–varianza alcanza un ratio de Sharpe de 0,867 y la de convicción de 0,836, frente a 0,314 del mercado y 0,217 de la cartera equiponderada; en términos de retorno, la adaptativa de media–varianza rinde 0,406 contra 0,125 del mercado. En el año bajista de 2022 —el más exigente de la muestra— el comportamiento es revelador: mientras el mercado pierde de forma pronunciada, con un ratio de Sharpe

de $-0{,}884$ y un retorno de $-0{,}167$, y los benchmarks históricos también incurren en pérdidas, las estrategias informacionales resisten mucho mejor. Aquí destaca la estática de convicción, que registra el mayor ratio de Sharpe del año (1,046) y el mayor retorno (0,339); las otras dos estáticas también resisten en terreno claramente positivo —0,600 y 0,628 en ratio de Sharpe—, mientras que entre las adaptativas solo tres se mantienen en terreno positivo —las de convicción y de media–varianza, ambas con 0,208, y la principal de máximo Sharpe, con 0,118—, en tanto la robusta ($-0{,}097$) y la de mínima varianza ($-0{,}429$) terminan en terreno negativo, aunque muy por encima del mercado en ambos casos. Estos resultados son consistentes con la interpretación del desacuerdo como una penalización informacional que reduce la exposición a activos con señales más frágiles durante regímenes de estrés (Carlin et al., 2014; Diether et al., 2002).

En los regímenes alcistas el panorama se invierte parcialmente. En 2021 y 2023 las estrategias adaptativas vuelven a liderar —las de media–varianza y de convicción alcanzan ratios de Sharpe superiores a 2,1 en 2021 y a 1,8 en 2023— y en 2025 casi todas las informacionales vuelven a superar al mercado, encabezadas por la estática de convicción (0,920) y la adaptativa de convicción (0,725). En cambio, en el año de mayor euforia bursátil (2024) el índice de mercado sube de forma tan generalizada que resulta difícil de batir, con un ratio de Sharpe de 1,702, y las estrategias informacionales, aunque competitivas, quedan algo por detrás. El cuadro de conjunto es, por tanto, que el aporte de estas estrategias no es uniforme entre regímenes, sino que se concentra en los años difíciles —el shock pandémico de 2020 y el mercado bajista de 2022— y en las fases de recuperación, donde ponderar los activos según su consenso y su desacuerdo les permite superar tanto al mercado como a los benchmarks históricos. Entre ellas, las variantes de convicción y de media–varianza son las más consistentes a lo largo de los seis años, mientras que la estática de convicción es la que mejor protege en el peor año, 2022.

Por último, la Tabla 14 caracteriza la rotación y la concentración de las carteras, dimensiones relevantes para juzgar la viabilidad de cada estrategia más allá del desempeño bruto. Las estrategias difieren de forma sustantiva en su comportamiento operativo. Sin contar la cartera equiponderada, la adaptativa de mínima varianza es la más diversificada y de menor rotación, con una rotación de 0,242, un número efectivo de 34,3 activos y un peso máximo de 0,066,

Tabla 14: Rotación y concentración de las carteras en la configuración ganadora.

Estrategia	Rotación	HHI	*N* efectivo	Peso máx.	*N* activos
S&P 500	—	—	—	—	—
Pesos iguales	0,000	0,022	46,0	0,022	46,0
Hist. mínima varianza	0,658	0,126	9,2	0,205	14,7
Hist. máximo Sharpe	0,949	0,289	4,3	0,389	6,8
Estática principal	0,490	0,072	14,5	0,132	22,4
Estática robusta	0,556	0,081	13,1	0,169	23,2
Estática convicción	0,594	0,134	7,9	0,250	22,4
Adaptativa principal (máx. Sharpe)	0,668	0,078	13,3	0,146	22,5
Adaptativa principal (mín. var.)	0,242	0,029	34,3	0,066	46,0
Adaptativa principal (media–var.)	0,780	0,161	6,9	0,248	11,0
Adaptativa robusta	0,818	0,164	7,6	0,303	20,7
Adaptativa convicción	0,756	0,132	8,0	0,236	22,5

Nota: la rotación es la fracción media de la cartera recompuesta en cada rebalanceo; el índice de Herfindahl–Hirschman (HHI) y el número efectivo de activos (N efectivo $= 1/\text{HHI}$) miden la concentración; N activos es el número medio de posiciones con peso no nulo. El índice S&P 500 es una estrategia pasiva y no admite estas medidas.

pues reparte el capital entre los activos de menor desacuerdo. En el extremo opuesto, el máximo Sharpe histórico es la más concentrada, con un número efectivo de 4,3 activos y un peso máximo de 0,389, y la de mayor rotación (0,949), lo que ilustra la inestabilidad de las soluciones de media–varianza clásicas. Las estrategias informacionales de máximo Sharpe y de convicción ocupan posiciones intermedias, con números efectivos de entre 7 y 14 activos y rotaciones de entre 0,49 y 0,82, y las adaptativas tienden a rotar algo más que sus contrapartes estáticas. Esta diferencia de rotación es pertinente porque el valor económico bruto del consenso debe ponderarse frente a los costos de transacción que su implementación conlleva (Barber et al., 2001): una rotación más alta erosiona el desempeño neto, de modo que la menor rotación de las estrategias estáticas y, sobre todo, de la adaptativa de mínima varianza constituye una ventaja que el desempeño bruto no refleja directamente. La cuantificación del impacto de los costos de transacción queda fuera del alcance de este trabajo, en el que las métricas se reportan brutas, y se señala como una extensión natural.

En síntesis, la evidencia respalda dos conclusiones centrales. Primero, separar el consenso del desacuerdo y tratar este último como riesgo informacional permite construir carteras que superan a los benchmarks clásicos en el pipeline que preserva la señal predictiva; esta ventaja frente a la referencia pasiva se sostiene incluso en 2022, el año más adverso de la muestra, en el que todas las estrategias informacionales resisten mejor que el índice. Segundo, la agregación adaptativa mediante aprendizaje secuencial no domina de forma uniforme a la agregación

estática, pero sí la iguala o supera en el horizonte de 120 días, donde alcanza su mejor desempeño relativo; esa ventaja se acumula a lo largo del ciclo, apoyada en los años de recuperación y de mercado alcista, y no en el episodio de estrés agudo de 2022, en el que la regla fija resiste mejor. Este resultado es importante porque la agregación estática constituye un rival exigente y no una referencia débil. El análisis de sensibilidad de la sección siguiente examina en qué medida estas conclusiones son robustas a la elección de hiperparámetros del meta-agente, al esquema de entrenamiento y a las restantes configuraciones experimentales.

5 Análisis de sensibilidad

La Sección 4 caracterizó el desempeño del sistema bajo su configuración principal. Esta sección examina hasta qué punto esas conclusiones dependen de decisiones de diseño, a través de tres ejercicios de sensibilidad ordenados desde lo más interno del meta-agente hacia lo más próximo a la construcción de las señales: la elección de hiperparámetros del meta-agente, el esquema temporal con que se entrena la política y la escala en que se expresa el consenso del crowd al construir los portafolios. Conviene precisar de antemano qué componentes afecta cada decisión. Los benchmarks de mercado, equiponderado e históricos no utilizan las señales del crowd y, por lo tanto, permanecen idénticos en todos los ejercicios; las estrategias estáticas tampoco dependen del meta-agente, de modo que son invariantes frente a la elección de hiperparámetros y al esquema de entrenamiento, aunque sí responden a la escala con que se construye el consenso. Por esta razón, las tablas de cada ejercicio reportan únicamente las filas que efectivamente varían e indican en sus notas los valores de las estrategias que permanecen como referencia. El análisis se concentra en el pipeline de retornos crudos, que domina de manera uniforme al de retornos tratados, y reporta en todos los casos tanto el agregado sobre los tres horizontes como el horizonte focal de 120 días, mediante el ratio de Sharpe y su equivalencia geométrica fuera de muestra.

5.1 Sensibilidad a los hiperparámetros del meta-agente

La política del meta-agente depende de un conjunto de hiperparámetros que gobiernan el presupuesto de entrenamiento, la frecuencia y la intensidad de la recolección de experien-

cia, el factor de descuento, la exploración durante la recolección y el modo de normalización de la recompensa. Para evaluar la estabilidad de las conclusiones frente a estas elecciones se consideran doce configuraciones, organizadas en cinco familias. La familia A varía el número de pases por evento de recolección manteniendo el resto de los valores del caso base; la familia B reduce el intervalo entre eventos de recolección; la familia C amplía el número de rondas de refinamiento; y las familias D y E combinan un factor de descuento menor con una exploración más intensa durante la recolección, contrastando además los dos modos de normalización de la recompensa: la reescalada min–max por horizonte, que mapea la recompensa al intervalo $[0, 1]$ a partir de los valores mínimo y máximo observados en el buffer al término de la fase de exploración, y un esquema de recorte y escala que opera con parámetros fijados de antemano. Este último primero recorta la recompensa cruda a un intervalo simétrico —entre -3 y 3 en la implementación—, con lo que acota la influencia de los valores atípicos, y a continuación la divide por una constante que crece con el horizonte —1,0 para 20 días, 1,5 para 60 y 2,0 para 120—, de manera que los tres horizontes queden en una escala comparable; a diferencia de la reescalada min–max, conserva el signo y el cero de la recompensa en lugar de comprimirla al intervalo $[0, 1]$. La Tabla 15 detalla las doce configuraciones.

Tabla 15: Configuraciones de hiperparámetros del meta-agente evaluadas en el análisis de sensibilidad.

Caso	Rondas	Recolección	Pases	Normalización	γ	ε_{rec}
A_1	400	100	1	min–max	0,98	0,30
A_2	400	100	2	min–max	0,98	0,30
A_3	400	100	3	min–max	0,98	0,30
B_1	400	50	1	min–max	0,98	0,30
B_2	400	50	2	min–max	0,98	0,30
B_3	400	50	3	min–max	0,98	0,30
C_1	800	100	1	min–max	0,98	0,30
C_2	800	100	2	min–max	0,98	0,30
D_1	600	100	2	min–max	0,95	0,40
D_2	600	100	2	recorte-escala	0,95	0,40
E_1	800	100	2	min–max	0,95	0,40
E_2	800	100	2	recorte-escala	0,95	0,40

Nota: El caso A_1 corresponde a la configuración principal empleada en la Sección 4.

Las Tablas 16 y 17 reportan el desempeño de las cinco estrategias adaptativas bajo cada configuración, en el agregado sobre horizontes y en el horizonte de 120 días respectivamente.

La lectura conjunta de las Tablas 16 y 17 arroja un patrón nítido de robustez. A lo largo de las doce configuraciones, el ratio de Sharpe agregado de cada estrategia adaptativa se man-

Tabla 16: Sensibilidad a los hiperparámetros del meta-agente: ratio de Sharpe agregado sobre horizontes en el pipeline RAW, con la equivalencia geométrica del retorno entre paréntesis.

Caso	P/máx. Sharpe	P/mín. var.	P/media–var.	Robusta	Convicción
A_1	0,891 (0,220)	0,608 (0,109)	0,796 (0,256)	0,785 (0,213)	0,862 (0,252)
A_2	0,829 (0,205)	0,621 (0,111)	0,705 (0,233)	0,749 (0,197)	0,777 (0,229)
A_3	0,898 (0,221)	0,642 (0,115)	0,762 (0,255)	0,797 (0,208)	0,862 (0,253)
B_1	0,779 (0,189)	0,590 (0,106)	0,654 (0,206)	0,713 (0,176)	0,720 (0,206)
B_2	0,793 (0,190)	0,589 (0,105)	0,675 (0,210)	0,711 (0,175)	0,707 (0,205)
B_3	0,791 (0,192)	0,589 (0,105)	0,662 (0,208)	0,751 (0,187)	0,733 (0,209)
C_1	0,889 (0,220)	0,586 (0,107)	0,806 (0,264)	0,736 (0,202)	0,866 (0,259)
C_2	0,852 (0,209)	0,594 (0,107)	0,702 (0,236)	0,736 (0,191)	0,814 (0,236)
D_1	0,800 (0,195)	0,601 (0,107)	0,793 (0,254)	0,719 (0,186)	0,761 (0,218)
D_2	0,866 (0,216)	0,593 (0,108)	0,826 (0,257)	0,709 (0,219)	0,846 (0,254)
E_1	0,835 (0,206)	0,597 (0,107)	0,793 (0,264)	0,682 (0,185)	0,801 (0,234)
E_2	0,844 (0,202)	0,539 (0,100)	0,788 (0,228)	0,736 (0,215)	0,851 (0,242)

Nota: P/máx. Sharpe, P/mín. var. y P/media–var. denotan la estrategia adaptativa principal optimizada por máximo Sharpe, mínima varianza y media–varianza, respectivamente; Robusta y Convicción son las variantes adaptativas correspondientes. Cada celda reporta el ratio de Sharpe y, entre paréntesis, la equivalencia geométrica. Como referencia invariante, en agregado el S&P 500 alcanza un Sharpe de 0,791 (0,135) y las estrategias estáticas principal, robusta y de convicción tienen un valor de 0,951 (0,223), 0,983 (0,227) y 0,957 (0,264), respectivamente.

tiene dentro de bandas estrechas: la principal de máximo Sharpe oscila entre 0,779 y 0,898, la de convicción entre 0,707 y 0,866, la de media–varianza entre 0,654 y 0,826, la robusta entre 0,682 y 0,797 y la de mínima varianza entre 0,539 y 0,642. Las conclusiones cualitativas centrales de la Sección 4 se mantienen al cambiar de configuración: las variantes de convicción y de media–varianza continúan concentrando el mejor desempeño en retorno, la estrategia de mínima varianza permanece como la alternativa más conservadora, y el horizonte de 120 días sigue siendo el tramo donde la capa adaptativa muestra su mayor aporte relativo. La robustez no implica que todas las configuraciones superen siempre a la exposición pasiva al mercado en el agregado, sino que el patrón principal de desempeño se conserva especialmente en las estrategias económicamente más relevantes. El caso principal A_1 resulta representativo y no un óptimo seleccionado a conveniencia: en agregado encabeza la estrategia de máximo Sharpe (0,891), solo superada por A_3 (0,898), y comparte con A_3 el primer lugar en la de convicción (0,862); en el horizonte de 120 días lidera tanto la variante de convicción (1,157) como la de media–varianza (1,129), las dos que sostienen el análisis económico de la Sección 4.4. La familia B, que reduce a la mitad el intervalo entre eventos de recolección, es la que más se rezaga en las estrategias de máximo Sharpe y de convicción, lo que sugiere que re-colectar experiencia con demasiada frecuencia sobre un presupuesto de rondas fijo no aporta y puede introducir cierta inestabilidad; en cambio, ampliar el número de rondas o modificar el factor de descuento

Tabla 17: Sensibilidad a los hiperparámetros del meta-agente: ratio de Sharpe en el horizonte de 120 días en el pipeline RAW, con la equivalencia geométrica del retorno entre paréntesis.

Caso	P/máx. Sharpe	P/mín. var.	P/media–var.	Robusta	Convicción
A_1	1,038 (0,248)	0,609 (0,107)	1,129 (0,339)	0,876 (0,225)	1,157 (0,324)
A_2	0,973 (0,226)	0,671 (0,115)	1,025 (0,312)	0,885 (0,210)	1,051 (0,294)
A_3	1,036 (0,238)	0,695 (0,119)	0,946 (0,304)	0,919 (0,215)	1,118 (0,308)
B_1	0,922 (0,218)	0,631 (0,113)	0,931 (0,292)	0,917 (0,212)	0,974 (0,279)
B_2	0,924 (0,215)	0,618 (0,111)	0,977 (0,293)	0,872 (0,201)	0,920 (0,269)
B_3	0,913 (0,219)	0,628 (0,113)	0,932 (0,289)	0,914 (0,219)	0,954 (0,272)
C_1	0,998 (0,233)	0,633 (0,111)	0,991 (0,301)	0,644 (0,167)	1,036 (0,300)
C_2	1,065 (0,249)	0,669 (0,117)	0,981 (0,313)	0,945 (0,229)	1,150 (0,313)
D_1	0,982 (0,227)	0,676 (0,116)	0,981 (0,292)	0,956 (0,229)	1,023 (0,280)
D_2	1,003 (0,242)	0,652 (0,116)	1,078 (0,314)	0,828 (0,275)	1,048 (0,303)
E_1	1,041 (0,249)	0,649 (0,115)	0,986 (0,311)	0,777 (0,202)	1,125 (0,322)
E_2	0,922 (0,217)	0,628 (0,112)	0,978 (0,259)	0,774 (0,260)	0,971 (0,267)

Nota: convenciones de columnas como en la Tabla 16. Como referencia invariante, en el horizonte de 120 días el S&P 500 alcanza un Sharpe de 0,791 (0,135) y las estrategias estáticas principal, robusta y de convicción tienen un valor de 1,002 (0,233), 1,149 (0,258) y 1,109 (0,303).

y la exploración produce resultados competitivos pero no sistemáticamente superiores a los del caso base. Esta estabilidad es coherente con que la corrección del sesgo de sobreestimación que introduce la arquitectura Double DQN favorece políticas menos sensibles al ruido de estimación (van Hasselt et al., 2016), y respalda la decisión de adoptar A_1 como configuración principal.

El diseño de las familias D y E permite, además, aislar el efecto del modo de normalización de la recompensa, puesto que D_2 y E_2 replican exactamente a D_1 y E_1 salvo por sustituir la reescalada min–max por el esquema de recorte y escala. El contraste limpio surge de comparar esos pares en la Tabla 16: de D_1 a D_2 y de E_1 a E_2, el esquema de recorte y escala tiende a favorecer ligeramente las estrategias de máximo Sharpe y de convicción —la de convicción sube de 0,761 a 0,846 entre D_1 y D_2, y de 0,801 a 0,851 entre E_1 y E_2— con un efecto mixto y de segundo orden sobre las variantes de mínima varianza y robusta. El modo de normalización es, por tanto, una decisión de implementación que altera la magnitud e incluso puede modificar marginalmente el orden relativo entre algunas estrategias, pero no cambia las conclusiones cualitativas del ejercicio: las variantes de máximo Sharpe, media–varianza y convicción siguen concentrando el desempeño más relevante, mientras que la mínima varianza conserva un perfil más defensivo.

5.2 Esquema de entrenamiento: ventana deslizante frente a historia expansiva

El segundo ejercicio contrasta dos formas de construir la muestra de entrenamiento del meta-agente. El esquema principal, denotado 4ENT, entrena la política de cada año de evaluación con las predicciones fuera de muestra de los cuatro años inmediatamente previos, manteniendo una ventana deslizante de longitud fija. El esquema alternativo, denotado HIST, reproduce en cambio la lógica de ventana de entrenamiento expansiva habitual en parte de la literatura de aprendizaje automático aplicada a la predicción de retornos (Gu et al., 2020), acumulando toda la historia fuera de muestra disponible hasta cada año de evaluación. La comparación permite discernir si conviene retener toda la experiencia pasada o privilegiar la más reciente. La Tabla 18 reporta ambos esquemas para el caso A_1 en el pipeline de retornos crudos.

Tabla 18: Sensibilidad al esquema de entrenamiento: ventana deslizante de cuatro períodos (4ENT) frente a historia expansiva (HIST), caso A_1 en el pipeline RAW. Cada celda reporta el ratio de Sharpe y, entre paréntesis, la equivalencia geométrica del retorno.

	Agregado		120 días	
Estrategia	4ENT	HIST	4ENT	HIST
Adaptativa principal (máx. Sharpe)	0,891 (0,220)	0,852 (0,205)	1,038 (0,248)	0,892 (0,204)
Adaptativa principal (mín. var.)	0,608 (0,109)	0,586 (0,107)	0,609 (0,107)	0,632 (0,112)
Adaptativa principal (media–var.)	0,796 (0,256)	0,749 (0,243)	1,129 (0,339)	0,884 (0,268)
Adaptativa robusta	0,785 (0,213)	0,741 (0,197)	0,876 (0,225)	0,573 (0,142)
Adaptativa convicción	0,862 (0,252)	0,776 (0,223)	1,157 (0,324)	0,926 (0,253)

Nota: los benchmarks y las estrategias estáticas no dependen del esquema de entrenamiento y conservan los valores de referencia indicados en la Tabla 16.

El esquema de ventana deslizante domina al de historia expansiva de manera bastante general. En agregado, 4ENT supera a HIST en las cinco estrategias adaptativas —por ejemplo, 0,891 frente a 0,852 en la de máximo Sharpe y 0,862 frente a 0,776 en la de convicción—, y la ventaja se acentúa en el horizonte de 120 días, donde la variante de convicción pasa de 0,926 bajo historia expansiva a 1,157 bajo ventana deslizante y la de media–varianza de 0,884 a 1,129. Extendido al conjunto de las doce configuraciones, el patrón se sostiene de forma robusta en la estrategia de convicción, donde 4ENT supera a HIST en el apartado agregado en once de los doce casos respecto al ratio de Sharpe y de forma total en el caso de retornos. En el horizonte de 120 días, la estrategia de convicción bajo 4ENT supera a HIST en todos los casos al considerar conjuntamente ratio de Sharpe y equivalencia geométrica; la única excepción sistemática es la estrategia de mínima varianza, cuyo desempeño es prácticamente indiferente

al esquema e incluso levemente superior bajo historia expansiva en el horizonte focal (0,632 frente a 0,609). La interpretación es económica más que técnica: en un entorno no estacionario, los regímenes antiguos pierden relevancia para anticipar la calidad relativa de los modelos, de modo que acumular indefinidamente la historia diluye la señal en lugar de reforzarla, y una ventana de horizonte acotado captura mejor las condiciones vigentes.

5.3 Escala del consenso: homogeneización frente a escala literal

El tercer ejercicio aborda una decisión de mayor calado conceptual: la escala en que se expresa el consenso al construir los portafolios. La convención operativa del estudio es homogeneizada, en el sentido de que tanto el consenso como el desacuerdo se expresan en una misma escala estandarizada, de modo que el ratio de Sharpe informacional compara dispersión relativa del crowd y no magnitudes de retorno. La alternativa literal mantiene el desacuerdo homogeneizado, pero expresa el consenso en escala de retorno esperado; con ello el numerador del problema de máximo Sharpe recupera unidades de retorno mientras el denominador conserva el significado de riesgo informacional. Esta distinción se traslada a dos puntos concretos del procedimiento. En el optimizador, bajo la escala homogeneizada el vector de retornos esperados proviene del consenso estandarizado y el problema de máximo Sharpe se resuelve sin descontar la tasa libre de riesgo en el numerador, ya que el cero representa la señal neutral; bajo la escala literal, ese vector se toma del consenso en retorno esperado y sí se descuenta la tasa libre de riesgo real, mientras que la matriz de riesgo —diagonal y construida a partir del desacuerdo homogeneizado— permanece idéntica, de modo que el cambio recae sobre el numerador del cociente que se optimiza y, por esa vía, sobre los pesos resultantes de la cartera. La recompensa del meta-agente, en cambio, no cambia de forma: en ambas escalas es el ratio de Sharpe realizado de la cartera resultante, calculado sobre los retornos diarios efectivos y la tasa libre de riesgo real fuera de muestra; lo que difiere es la cartera que esa recompensa evalúa, puesto que la escala altera el portafolio que el optimizador selecciona para cada acción del agente. A diferencia de los dos ejercicios anteriores, esta decisión afecta a todas las estrategias informacionales —estáticas y adaptativas—, ya que modifica el insumo direccional común a todas ellas; los benchmarks, que no emplean el consenso del crowd, permanecen invariantes.

La Tabla 19 compara ambas escalas para el caso A_1 en el pipeline de retornos crudos.

Tabla 19: Sensibilidad a la escala del consenso: escala homogeneizada (convención operativa) frente a escala literal, caso A_1 en el pipeline RAW. Cada celda reporta el ratio de Sharpe y, entre paréntesis, la equivalencia geométrica.

	Agregado		120 días	
Estrategia	Homogénea	Literal	Homogénea	Literal
Estática principal	0,951 (0,223)	0,908 (0,197)	1,002 (0,233)	0,996 (0,211)
Estática robusta	0,983 (0,227)	0,899 (0,189)	1,149 (0,258)	0,979 (0,194)
Estática convicción	0,957 (0,264)	0,950 (0,253)	1,109 (0,303)	1,056 (0,268)
Adaptativa principal (máx. Sharpe)	0,891 (0,220)	0,851 (0,211)	1,038 (0,248)	1,144 (0,266)
Adaptativa principal (mín. var.)	0,608 (0,109)	0,567 (0,105)	0,609 (0,107)	0,548 (0,098)
Adaptativa principal (media–var.)	0,796 (0,256)	0,787 (0,159)	1,129 (0,339)	0,884 (0,183)
Adaptativa robusta	0,785 (0,213)	0,756 (0,194)	0,876 (0,225)	0,959 (0,219)
Adaptativa convicción	0,862 (0,252)	0,871 (0,254)	1,157 (0,324)	1,246 (0,339)

Nota: los benchmarks no emplean el consenso del crowd y permanecen invariantes ante la escala, con los valores de referencia indicados en la Tabla 16.

El cuadro que emerge es matizado y respalda la elección de la escala homogeneizada como convención operativa. En agregado, la escala homogeneizada es al menos tan buena o mejor que la literal en la mayoría de las estrategias: las tres estáticas y las adaptativas de máximo Sharpe, mínima varianza, media–varianza y robusta ceden algo de desempeño bajo la escala literal, mientras que solo la adaptativa de convicción se mantiene prácticamente igual (0,871 frente a 0,862). En el horizonte de 120 días el panorama es más heterogéneo: la escala literal mejora a las adaptativas de máximo Sharpe (de 1,038 a 1,144), de convicción (de 1,157 a 1,246) y robusta (de 0,876 a 0,959), pero deteriora con claridad a la de media–varianza, cuya equivalencia geométrica cae de 0,339 a 0,183, y a la de mínima varianza. La escala homogeneizada ofrece así resultados más consistentes entre estrategias y horizontes, lo que justifica su adopción como convención principal, en tanto la escala literal queda como una alternativa de contraste. Lo relevante para la validez del estudio es que las conclusiones cualitativas se conservan bajo ambas escalas: las estrategias informacionales mantienen una ventaja sustantiva frente a los benchmarks en las configuraciones centrales, y la variante de convicción conserva su carácter de mejor desempeño en el horizonte focal bajo ambas escalas.

En síntesis, los tres ejercicios apuntan en la misma dirección: el desempeño documentado en la Sección 4 es robusto a la elección de hiperparámetros del meta-agente, se beneficia de una ventana de entrenamiento deslizante frente a una expansiva y se sostiene cualitativamente bajo distintas formas de procesar las señales. Esto refuerza la interpretación de que el

aporte proviene, primero, del contenido informacional del crowd expresado en consenso y desacuerdo, y segundo, de la reponderación dinámica que introduce el meta-agente en horizontes y regímenes específicos, más que de una sintonización particular del sistema. Un eje de sensibilidad adicional, no abordado aquí y planteado como extensión natural, es la configuración interna del filtrado wavelet del pipeline de retornos tratados —la familia wavelet, el nivel de descomposición y la regla de reconstrucción—, cuya exploración sistemática podría precisar las condiciones bajo las cuales el preprocesamiento del objetivo ayuda o perjudica (Chan Phooi M'ng & Mehralizadeh, 2016; Nobre & Neves, 2019).

6 Conclusiones

Este trabajo abordó la construcción de portafolios a partir de un crowd artificial de modelos predictivos heterogéneos, proponiendo tratar de forma conjunta tres limitaciones habituales de los enfoques existentes: la reducción del desacuerdo entre fuentes a ruido descartable, el uso de reglas de agregación estáticas y simétricas, y la estimación del riesgo a partir de la historia de retornos de los activos antes que de la fragilidad de la información con que se decide. La propuesta descansa en tres reinterpretaciones. Primero, el desacuerdo entre los modelos del crowd se concibe como riesgo informacional, separando explícitamente el consenso —la señal direccional agregada que provee el vector de retornos esperados— de la dispersión entre fuentes —que provee una matriz de riesgo diagonal—. Segundo, un meta-agente basado en Double Deep Q-Learning reemplaza la regla de agregación fija por una dependiente del estado, aprendiendo a reponderar a los miembros del crowd según el régimen de mercado. Tercero, el problema de Markowitz se reinterpreta en clave informacional, de modo que sus dos insumos no son momentos estimados de los retornos, sino el consenso y el desacuerdo del crowd, lo que evita la estimación directa de covarianzas de retornos.

La evaluación empírica sobre cuarenta y seis acciones estadounidenses, tres horizontes de inversión y un esquema walk-forward con seis años de evaluación arrojó un conjunto coherente de resultados. En el plano predictivo, la heterogeneidad del crowd se manifestó como una división de roles: los modelos de árboles y las redes neuronales concentraron la capacidad direccional, especialmente en el horizonte de ciento veinte días, mientras que los híbridos de

volatilidad y las reglas de promedio móvil exhibieron una precisión de signo sistemáticamente menor, lo que sugiere un rol complementario antes que direccional. En el plano económico, y en el pipeline que preserva la señal predictiva en escala cruda, las estrategias informacionales mostraron un desempeño superior al de los benchmarks clásicos en la mayoría de las especificaciones relevantes, especialmente en las estáticas y en las variantes adaptativas de máximo Sharpe, media–varianza y convicción. Un hallazgo que merece subrayarse es que la agregación estática constituyó un rival exigente, exhibiendo los ratios de Sharpe agregados más altos, en línea con la robustez que la literatura de combinación de pronósticos atribuye a las reglas simples (Timmermann, 2006); el aporte de la capa adaptativa no fue uniforme, sino que se concentró en el horizonte de ciento veinte días, donde la variante de convicción alcanzó el mayor ratio de Sharpe del horizonte, por un margen estrecho sobre la mejor especificación estática. Esa ventaja se construyó a lo largo del ciclo de evaluación, apoyada en los años de recuperación y de mercado alcista, y no de manera homogénea año a año: en el episodio de estrés agudo de 2022, la agregación estática resistió mejor que la adaptativa. En contraste, el tratamiento del objetivo mediante filtrado wavelet y winsorización deterioró el desempeño económico de forma generalizada, al erosionar la señal direccional de la que dependen las estrategias de máximo Sharpe y de convicción.

Los análisis de sensibilidad respaldaron la solidez de estas conclusiones. El desempeño resultó robusto a la elección de los hiperparámetros del meta-agente, de modo que la configuración principal se mostró representativa y no un óptimo seleccionado a conveniencia; el esquema de entrenamiento con ventana deslizante superó de manera general al de historia expansiva, consistente con la pérdida de relevancia de los regímenes antiguos en un entorno no estacionario; y las conclusiones cualitativas centrales se conservaron bajo las dos escalas consideradas para expresar el consenso.

En conjunto, la evidencia sostiene dos conclusiones centrales. La primera es que separar el consenso del desacuerdo y tratar este último como riesgo informacional permite construir carteras que superan a los benchmarks clásicos en el pipeline que preserva la señal predictiva; esta ventaja frente a la referencia pasiva se sostiene incluso en el año de mayor tensión de la muestra, en el que todas las estrategias informacionales —estáticas y adaptativas— resis-

tieron mejor que el índice. La segunda es que la agregación adaptativa mediante aprendizaje secuencial no domina de manera uniforme a la agregación estática, pero la iguala o supera en el horizonte de ciento veinte días, donde alcanza su mejor desempeño relativo; esa ventaja se acumula a lo largo del ciclo, apoyada en los años de recuperación y de mercado alcista, y no en el episodio de estrés agudo, en el que la regla fija se mostró más resistente. El carácter acotado de esa ventaja adquiere mayor valor precisamente porque la agregación estática es un rival exigente y no una referencia débil.

6.1 Limitaciones y trabajo futuro

Los resultados deben interpretarse a la luz de varias limitaciones, cada una de las cuales sugiere una extensión natural. En primer lugar, todas las métricas se reportan brutas, sin descontar costos de transacción; esta omisión es particularmente relevante porque las estrategias adaptativas de mejor desempeño exhiben una rotación más alta que sus contrapartes estáticas, de modo que su ventaja bruta podría reducirse —o incluso revertirse— una vez incorporados los costos de implementación (Barber et al., 2001). En segundo lugar, el universo empírico se restringe a cuarenta y seis acciones de un único mercado, por lo que la generalización a otros mercados, clases de activo o universos de mayor tamaño queda por establecer. Además, la exigencia de trabajar con un panel balanceado de activos con información completa puede inducir un sesgo de selección hacia empresas más líquidas, persistentes y con trayectorias observables a lo largo de todo el período. Esta decisión mejora la comparabilidad estadística entre activos y evita problemas derivados de datos faltantes, pero limita la extrapolación de los resultados a universos más amplios, no balanceados o sujetos a entrada y salida de activos. En tercer lugar, la matriz de riesgo informacional es diagonal por construcción, decisión que aporta estabilidad numérica y evita la estimación de covarianzas, pero que no captura la dependencia entre activos: el riesgo informacional aquí propuesto es relacional entre fuentes, no entre activos. En cuarto lugar, la cartera se restringe a posiciones largas y el espacio de acciones del meta-agente es discreto, lo que acota el conjunto de ponderaciones alcanzables. Asimismo, el análisis evalúa el desempeño económico de la política aprendida, pero no descompone de forma exhaustiva los mecanismos internos de decisión del meta-agente: queda pendiente estudiar qué modelos o

familias reciben mayor peso en cada régimen, qué variables del estado activan los cambios de política y si las acciones seleccionadas responden a patrones financieros interpretables, una caracterización que complementaría la evidencia de desempeño con una lectura más transparente del mecanismo adaptativo. Por último, la disponibilidad de fechas de decisión genuinamente fuera de muestra es estructuralmente escasa, lo que limita la complejidad de la política que el agente puede aprender, y el pipeline de tratamiento se evaluó bajo una única configuración de filtrado.

Estas limitaciones delinean una agenda de trabajo futuro. La extensión más inmediata consiste en incorporar costos de transacción de forma explícita, evaluando el desempeño neto y, eventualmente, penalizando la rotación dentro del propio problema de asignación. Una segunda línea consiste en ampliar el universo a otros mercados y clases de activo —incluidos los criptoactivos, donde arquitecturas de aprendizaje por refuerzo basadas en transformadores han mostrado resultados promisorios (Zepeda et al., 2026)—. Una tercera dirección consiste en enriquecer la estructura de riesgo informacional más allá de la diagonal, capturando la dependencia entre las señales de distintos activos, lo que conecta de forma natural con las formulaciones relacionales de asignación de riesgo basadas en la teoría de juegos cooperativos (Hiller, 2025b; Shalit, 2021). Asimismo, el carácter agnóstico del marco respecto del signo de la prima por desacuerdo sugiere explorar tratamientos condicionados al régimen, que alternen entre penalizar y recompensar la dispersión según las fricciones de mercado predominantes (Carlin et al., 2014; Miller, 1977). Por último, resulta natural explorar espacios de acción continuos y arquitecturas de aprendizaje alternativas, así como un barrido sistemático de la configuración del filtrado wavelet —familia, nivel de descomposición y regla de reconstrucción— con el fin de precisar las condiciones bajo las cuales el preprocesamiento del objetivo ayuda o perjudica (Chan Phooi M'ng & Mehralizadeh, 2016; Nobre & Neves, 2019).

En su conjunto, este trabajo muestra que, dentro del universo y protocolo empírico evaluado, tratar el desacuerdo entre un crowd de modelos como riesgo informacional y aprender a reponderar ese crowd de forma dependiente del estado constituye una vía viable para la construcción de portafolios informacionales. Su valor se manifiesta con mayor nitidez en el horizonte largo, donde el contenido económico de la señal es más alto, y cuando la agregación

adaptativa opera sobre señales direccionales preservadas en escala cruda.

Referencias

Alaminos, D., Salas, M. B., & Fernández-Gámez, M. Á. (2024). Hybrid genetic algorithms in agent-based artificial market model for simulating fan tokens trading. *Engineering Applications of Artificial Intelligence*, *131*, 107713. https://doi.org/10.1016/j.engappai.2023.107713

Auer, B. R., & Hiller, T. (2021). Cost gap, Shapley, or nucleolus allocation: Which is the best game-theoretic remedy for the low-risk anomaly? *Managerial and Decision Economics*, *42*(4), 876-884. https://doi.org/10.1002/mde.3279

Barber, B., Lehavy, R., McNichols, M., & Trueman, B. (2001). Can Investors Profit from the Prophets? Security Analyst Recommendations and Stock Returns. *The Journal of Finance*, *56*(2), 531-563. https://doi.org/10.1111/0022-1082.00336

Bhandari, H. N., Rimal, B., Pokhrel, N. R., Rimal, R., Dahal, K. R., & Khatri, R. K. (2022). Predicting stock market index using LSTM. *Machine Learning with Applications*, *9*, 100320. https://doi.org/10.1016/j.mlwa.2022.100320

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*(3), 307-327. https://doi.org/10.1016/0304-4076(86)90063-1

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*, 5-32. https://doi.org/10.1023/A:1010933404324

Carlin, B. I., Longstaff, F. A., & Matoba, K. (2014). Disagreement and asset prices. *Journal of Financial Economics*, *114*(2), 226-238. https://doi.org/10.1016/j.jfineco.2014.06.007

Chan Phooi M'ng, J., & Mehralizadeh, M. (2016). Forecasting East Asian Indices Futures via a Novel Hybrid of Wavelet-PCA Denoising and Artificial Neural Network Models. *PLOS ONE*, *11*(6), 1-29. https://doi.org/10.1371/journal.pone.0156338

Chen, H., Joslin, S., & Tran, N.-K. (2010). Affine Disagreement and Asset Pricing. *American Economic Review*, *100*(2), 522-26. https://doi.org/10.1257/aer.100.2.522

Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785-794. https://doi.org/10.1145/2939672.2939785

Chow, G., Jacquier, E., Kritzman, M., & Lowry, K. (1999). Optimal Portfolios in Good Times and Bad. *Financial Analysts Journal*, *55*(3), 65-73. https://doi.org/10.2469/faj.v55.n3.2273

Diether, K. B., Malloy, C. J., & Scherbina, A. (2002). Differences of Opinion and the Cross Section of Stock Returns. *The Journal of Finance*, *57*(5), 2113-2141. https://doi.org/10.1111/0022-1082.00490

Donoho, D. (1995). De-noising by soft-thresholding. *IEEE Transactions on Information Theory*, *41*(3), 613-627. https://doi.org/10.1109/18.382009

Fasth, T., Larsson, A., & Kalinina, M. (2016). Disagreement Constrained Action Selection in Participatory Portfolio Decision Analysis. *International Journal of Innovation, Management and Technology*, *7*(1), 1-7. https://doi.org/10.18178/ijimt.2016.7.1.636

Friedman, J. H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics*, *29*(5), 1189-1232. https://doi.org/10.1214/aos/1013203451

García-Medina, A., & Aguayo-Moreno, E. (2024). LSTM–GARCH Hybrid Model for the Prediction of Volatility in Cryptocurrency Portfolios. *Computational Economics*, *63*, 1511-1542. https://doi.org/10.1007/s10614-023-10373-8

Gu, S., Kelly, B., & Xiu, D. (2020). Empirical Asset Pricing via Machine Learning. *The Review of Financial Studies*, *33*(5), 2223-2273. https://doi.org/10.1093/rfs/hhaa009

Hiller, T. (2025a). Cooperative Game Theory of Hierarchies: One Approach to Solving the Low-Risk Puzzle? *Risks*, *13*(5). https://doi.org/10.3390/risks13050086

Hiller, T. (2025b). Weighted Shapley values and allocation of portfolio risk: one approach to solve the low-risk puzzle? *Theory and Decision*. https://doi.org/10.1007/s11238-025-10108-1

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, *9*(8), 1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735

Kamil, A. A., Mustafa, A., & Ibrahim, K. (2010). Maximum Downside Semi-Deviation Stochastic Programming for Portfolio Optimization Problem. *Journal of Modern Applied Statistical Methods*, *9*(2), 536-546. https://doi.org/10.22237/jmasm/1288585200

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: A Highly Efficient Gradient Boosting Decision Tree. *Advances in Neural Information Processing Systems*, *30*.

Kristjanpoller, W., Michell, K., Minutolo, M. C., & Dheeriya, P. (2021). Trading support system for portfolio construction using wisdom of artificial crowds and evolutionary computation. *Expert Systems with Applications*, *177*, 114943. https://doi.org/10.1016/j.eswa.2021.114943

Meeuwis, M., Parker, J. A., Schoar, A., & Simester, D. (2022). Belief Disagreement and Portfolio Choice. *The Journal of Finance*, *77*(6), 3191-3247. https://doi.org/10.1111/jofi.13179

Miller, E. M. (1977). RISK, UNCERTAINTY, AND DIVERGENCE OF OPINION. *The Journal of Finance*, *32*(4), 1151-1168. https://doi.org/10.1111/j.1540-6261.1977.tb03317.x

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., & Hassabis, D. (2015). Human-level control through deep reinforcement learning. *Nature*, *518*(7540), 529-533. https://doi.org/10.1038/nature14236

Nelson, D. B. (1991). Conditional Heteroskedasticity in Asset Returns: A New Approach. *Econometrica*, *59*(2), 347-370. https://doi.org/10.2307/2938260

Nobre, J., & Neves, R. F. (2019). Combining Principal Component Analysis, Discrete Wavelet Transform and XGBoost to trade in the financial markets. *Expert Systems with Applications*, *125*, 181-194. https://doi.org/10.1016/j.eswa.2019.01.083

Pandelara, D., Kristjanpoller, W., Michell, K., & Minutolo, M. C. (2022). A fuzzy regression causality approach to analyze relationship between electrical consumption and GDP. *Energy*, *239*, 122459. https://doi.org/10.1016/j.energy.2021.122459

Qian, Y., & Zhang, Y. (2025). Long-term forecasting in asset pricing: Machine learning models' sensitivity to macroeconomic shifts and firm-specific factors. *The North American Journal of Economics and Finance*, *78*, 102423. https://doi.org/10.1016/j.najef.2025.102423

Shahparast, H., Hamzeloo, S., & Safari, E. (2023). An incremental type-2 fuzzy classifier for stock trend prediction. *Expert Systems with Applications*, *212*, 118787. https://doi.org/10.1016/j.eswa.2022.118787

Shalit, H. (2020). The Shapley value of regression portfolios. *Journal of Asset Management*, *21*, 506-512. https://doi.org/10.1057/s41260-020-00175-0

Shalit, H. (2021). The Shapley value decomposition of optimal portfolios. *Annals of Finance*, *17*, 1-25. https://doi.org/10.1007/s10436-020-00380-2

Sikalo, M., Arnaut-Berilo, A., & Zaimovic, A. (2022). Efficient Asset Allocation: Application of Game Theory-Based Model for Superior Performance. *International Journal of Financial Studies*, *10*(1). https://doi.org/10.3390/ijfs10010020

Thakkar, A., & Chaudhari, K. (2024). Applicability of genetic algorithms for stock market prediction: A systematic survey of the last decade. *Computer Science Review*, *53*, 100652. https://doi.org/10.1016/j.cosrev.2024.100652

Timmermann, A. (2006). Chapter 4 Forecast Combinations. En G. Elliott, C. Granger & A. Timmermann (Eds.). Elsevier. https://doi.org/10.1016/S1574-0706(05)01004-9

van Hasselt, H., Guez, A., & Silver, D. (2016). Deep Reinforcement Learning with Double Q-Learning. *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16)*, 2094-2100.

Wang, J., & Chen, Z. (2023). Exploring Low-Risk Anomalies: A Dynamic CAPM Utilizing a Machine Learning Approach. *Mathematics*, *11*(14). https://doi.org/10.3390/math11143220

Yu, J. (2011). Disagreement and return predictability of stock portfolios. *Journal of Financial Economics*, *99*(1), 162-183. https://doi.org/10.1016/j.jfineco.2010.08.004

Zepeda, M., Kristjanpoller, W., & Minutolo, M. C. (2026). Deep transformer Q-learning based reinforcement learning for portfolio optimization of cryptocurrencies. *Applied Soft Computing*, *195*, 115042. https://doi.org/10.1016/j.asoc.2026.115042