

INTELIGENCIA ARTIFICIAL EN LOS MERCADOS FINANCIEROS: EVIDENCIA EMPÍRICA EN EL S&P 500

Iván Blanco

Universidad. CUNEF Universidad

Resumen

Este artículo analiza el papel de la inteligencia artificial en los mercados financieros, explorando su capacidad para mejorar la predicción de retornos y la construcción de carteras. A través de un caso práctico aplicado al S&P 500, se emplean técnicas de *machine learning* sobre un conjunto amplio de factores, mostrando que los modelos generan señales consistentes y rentabilidades superiores al índice de referencia, incluso tras considerar los costes de transacción. Los resultados ilustran cómo la inteligencia artificial puede integrarse en la gestión de activos institucional, combinando rigor académico con aplicabilidad práctica.

Palabras clave: inteligencia artificial, *machine learning*, gestión de activos, predicción de retornos, valoración de activos.

Abstract

This article examines the role of artificial intelligence in financial markets, focusing on its ability to improve return prediction and portfolio construction. Using a practical case study on the S&P 500, machine learning techniques are applied to a broad set of factors, showing that the models deliver consistent signals and outperform a benchmark even after accounting for transaction costs. The findings illustrate how AI can be integrated into institutional asset management, combining academic rigor with practical applicability.

Keywords: artificial intelligence, machine learning, asset management, return prediction, asset pricing.

JEL classification: C45, C55, G11, G12.

I. INTRODUCCIÓN

Se estima que entre un 60-80 por 100 del volumen de transacciones en los mercados de acciones de Estados Unidos se realiza mediante algoritmos y estrategias sistemáticas. Fondos cuantitativos que emplean inteligencia artificial (IA) gestionan carteras multimillonarias a nivel global. Firms como Renaissance Technologies, Two Sigma o Citadel han demostrado consistentemente que los modelos de IA pueden generar retornos superiores al mercado, incluso después de comisiones. Un ejemplo paradigmático es el Medallion Fund de Renaissance Technologies, que habría logrado un retorno anual compuesto de aproximadamente el 31-35 por 100 antes de comisiones durante más de tres décadas (Guo y Liu, 2024).

Esta revolución no es casual. La IA está redefiniendo la forma en que se genera información, se procesan grandes cantidades de datos y, por tanto, la manera en que se toman decisiones a nivel personal y empresarial. Los mercados financieros poseen dos características que los hacen especialmente sensibles a los efectos disruptivos de la IA: por una parte, la creciente cantidad de información que afecta a los precios de los activos financieros y; por otra, la enorme dificultad para predecir el comportamiento de dichos activos. En este contexto, cualquier pequeña mejora en la capacidad de predicción, por insignificante que parezca, puede tener consecuencias económicas relevantes. De ahí que, prácticamente desde los primeros avances en aprendizaje automático (*machine learning*, ML), estas técnicas se hayan ido introduciendo de manera acelerada, inicialmente en la industria de la gestión de activos y

fondos de cobertura (*hedge funds*) y, posteriormente, en el resto del sistema financiero.

Lo que comenzó como ventaja exclusiva de los *hedge funds* más sofisticados se está democratizando rápidamente. Hoy en día, gracias a la computación en la nube (*cloud*), las librerías y lenguajes de programación de código abierto y el uso de la IA como asistente en la programación, incluso inversores individuales pueden acceder a herramientas que hace una década costaban miles de dólares y estaban reservadas a equipos institucionales con costosos departamentos de analistas cuantitativos (*quants*). En esta nueva realidad, la ventaja competitiva ya no es únicamente el acceso a la tecnología, sino sobre todo la capacidad de adquirir, procesar y explotar fuentes de datos alternativos, desde imágenes satelitales hasta el “sentimiento” en redes

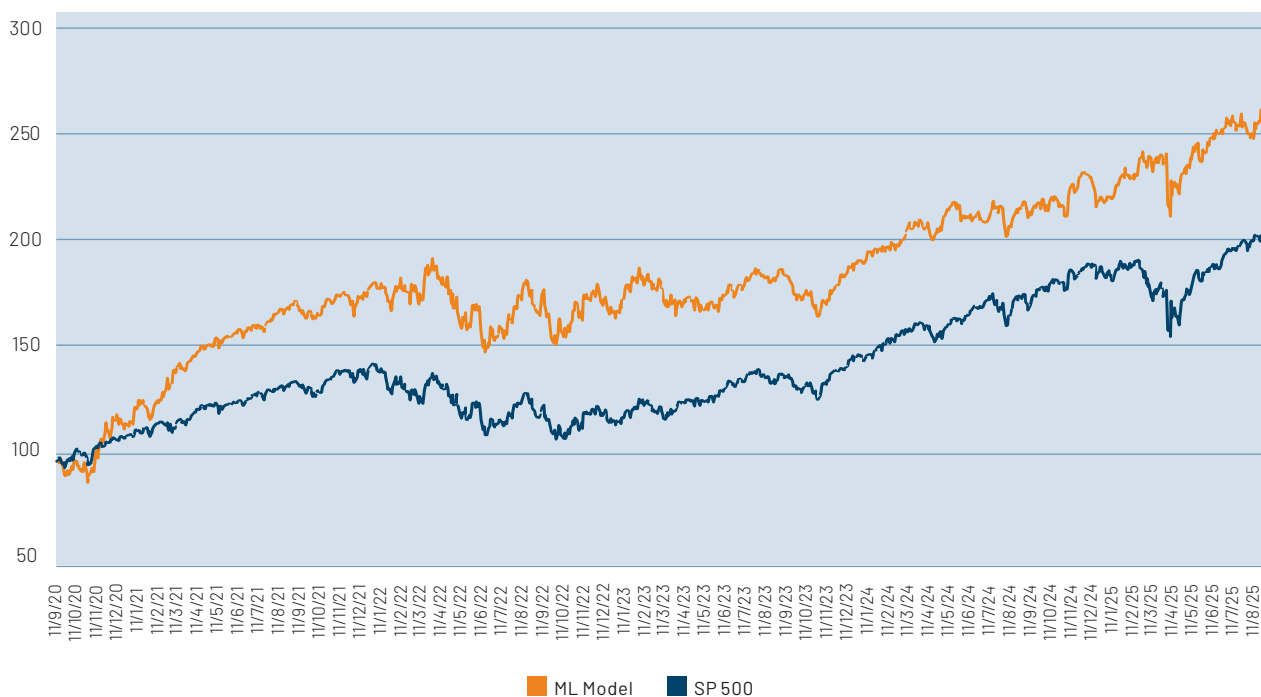
sociales, estos datos se han convertido en uno de los principales ejes de ventaja competitiva en la industria. Sin embargo, esta carrera hacia modelos cada vez más complejos genera una tensión fundamental: los algoritmos de la IA ofrecen mayor capacidad predictiva, pero su carácter de “caja negra” dificulta la interpretación, lo que plantea retos tanto para la gestión de riesgos como para el cumplimiento regulatorio.

En este artículo, además de repasar los fundamentos teóricos y el debate académico, se desarrolla un caso práctico diseñado para captar la atención del lector: una estrategia de inversión con *ML*, implementada sobre el universo del S&P 500, es decir, las empresas más líquidas y seguidas del mundo. El objetivo no es demostrar un ejercicio artificial en datos exóticos o ilíquidos, sino mostrar

GRÁFICO 1

RETORNOS ACUMULADOS DE LA ESTRATEGIA ML MODEL VS ÍNDICE S&P 500 FUERA DEL CONJUNTO DE DATOS DE ENTRENAMIENTO

Período desde 9/11/2020 hasta 09/11/2025



Fuente: Elaboración propia a partir de datos de FactSet.

que incluso en el segmento más competitivo del mercado pueden encontrarse oportunidades mediante la IA. El gráfico 1 ofrece un adelanto de los resultados. Entre septiembre de 2020 y septiembre de 2025, la estrategia basada en *ML* logra una rentabilidad anualizada del 21,5 por 100, frente al 16 por 100 del S&P 500, generando un alfa del 6,55 por 100 anualizado. Aunque la volatilidad de la estrategia es algo mayor a la del índice S&P 500 (*benchmark*), las ratios de Sharpe y Sortino superan con claridad a los del *benchmark*, y la máxima caída (*drawdown* máximo) resulta ligeramente inferior a la del índice como veremos en la sección cuatro del artículo. Esta evidencia preliminar, que se desarrolla en detalle más adelante, ilustra un mensaje central: la IA no solo es objeto de debate académico, sino una herramienta que puede traducirse en estrategias de inversión implementables por cualquier gestora con acceso a datos de calidad y recursos tecnológicos adecuados.

El objetivo de este artículo es guiar al lector desde la intuición económico-financiera necesaria para entender por qué la IA es relevante para los mercados financieros (sección segunda), explorando la pregunta fundamental sobre si realmente es posible predecir el mercado, introduciendo el concepto de inversión basada en factores (sección tercera), para después desarrollar una aplicación práctica con datos reales, la estrategia de *ML* sobre el S&P 500, que puede ser implementada por cualquier gestora de activos (Sección cuarta), y, finalmente, concluir con una breve discusión sobre los potenciales riesgos para los mercados si la IA no se usa con el criterio y la supervisión adecuados (sección quinta). En la medida de lo posible, se combinarán resultados académicos con ejemplos prácticos que puedan resultar útiles tanto para profesionales de la industria como para inversores, evitando tecnicismos innecesarios y centrándose en la intuición subyacente.

1. Machine learning & deep learning vs econometría

Antes de entrar en materia, se definirán brevemente algunos términos que habitualmente englobamos bajo el paraguas genérico de "inteligencia artificial", así como el concepto de "econometría", que usaremos a lo largo del artículo. Estas descrip-

ciones, sin ánimo de herir sensibilidades académicas o profesionales, son deliberadamente simplificadas para facilitar el entendimiento del artículo:

- La econometría busca entender las relaciones causales entre variables económicas mediante modelos estadísticos interpretables, priorizando la comprensión del "por qué" detrás de los fenómenos observados.
- El *machine learning* (*ML*), por su parte, se centra más en la capacidad predictiva que en la interpretabilidad, utilizando algoritmos que aprenden patrones complejos de los datos sin necesariamente explicar las relaciones subyacentes.
- El *deep learning* (*DL*), formalmente un subconjunto de técnicas de *machine learning*, emplea redes neuronales con múltiples capas que pueden descubrir representaciones cada vez más abstractas de los datos, siendo especialmente efectivo con información no estructurada como texto, imágenes o audio.

Dicho esto, a lo largo de este artículo, en muchas ocasiones se usará el término *machine learning*, *ML*, para referirse a modelos que no son de *deep learning*, obviando que estos son un subconjunto de los modelos de *machine learning*.

Para terminar de fijar las anteriores definiciones, se usará un ejemplo sencillo. Supongamos que queremos entender la relación entre los tipos de interés y el comportamiento de los precios de las empresas de pequeña capitalización en Estados Unidos (*small caps*). Un economista usaría, por ejemplo, una regresión lineal, o un modelo de diferencias en diferencias (*DiD*) para entender cómo afecta el tipo de interés al retorno de estas acciones. Un modelo de *ML*, en cambio, buscaría la combinación óptima de variables para predecir ese precio, priorizando la capacidad predictiva del modelo. Por su parte, un modelo de *DL* podría ir mucho más allá e incorporar información no estructurada: noticias financieras, sentimiento del mercado, imágenes por satélite de aparcamientos de centros comerciales o declaraciones de directivos, todo ello sin necesidad de especificar qué variables concretas debe considerar el modelo predictivo.

Esta última característica, la capacidad de trabajar sin especificar de antemano las variables a usar, es probablemente una de las diferencias fundamentales entre el ML tradicional y el DL.

2. Trading algorítmico vs inversión sistemática

La última aclaración pertinente antes de comenzar con el marco teórico conceptual es la distinción que haremos entre "trading algorítmico" e "inversión sistemática". Ambas se pueden englobar en lo que comúnmente se define como inversión cuantitativa, pero difieren, en general, en la forma en la que se ejecutan las órdenes en el mercado. A lo largo de este artículo nos referiremos al "trading algorítmico" cuando queramos enfatizar que las órdenes son ejecutadas directamente por un ordenador sin intervención humana en dicha ejecución. Por tanto, abusando del término, en general nos referiremos la inversión sistemática como a un conjunto más amplio de estrategias de inversión, basadas en señales, que pueden o no ser ejecutadas por un ordenador. En otras palabras, asumiremos que el trading algorítmico está incluido dentro de la definición de inversión sistemática, pero que esta abarca también la ejecución por parte de *traders* humanos siempre que se ciñan exclusivamente a seguir y ejecutar, de la mejor forma posible, las señales generadas por un sistema. En este artículo, como se verá en la sección cuarta, dicho sistema se apoyará en la IA.

II. MARCO CONCEPTUAL

1. IA y precios como predicciones

Una forma intuitiva de entender el impacto de la IA en la valoración de activos y, por ende, en cómo se toman las decisiones de inversión que afectan a los mercados, es partir de la idea central de que los precios contienen predicciones de los pagos futuros y de las preferencias de los inversores. En otras palabras, el precio de cualquier activo es un reflejo de los pagos futuros de ese activo, junto con su incertidumbre, descontados al presente teniendo en cuenta cómo valoran los inversores esos pagos futuros esperados. Esta intuición se recoge en la teoría moderna de valoración de activos y puede

formalizarse mediante una ecuación sencilla (Cochrane, 2005):

$$p_t = E_t [m_{t+1} + x_{t+1}], \quad [1]$$

donde:

p_t es el precio del activo hoy,

x_{t+1} son los pagos futuros (dividendos, cupones, precio de reventa),

m_{t+1} es el factor de descuento estocástico, que refleja cómo los inversores valoran recibir un pago en el futuro en lugar de hoy, y

$E_t [.]$ denota la expectativa condicional con respecto a toda la información disponible en el momento t .

Esta formulación resume de forma elegante la esencia de las finanzas: el precio actual es la predicción, bajo la información disponible hoy, del valor que los inversores asignarán a los pagos futuros. Dicho de otra manera, los precios no son números arbitrarios, sino que ya contienen la huella de las expectativas de los agentes económicos.

El reto surge al intentar llevar a la práctica esta idea. ¿Qué información se debe usar y cómo debe combinarse para formar una expectativa fiable? En un mundo con pocos datos y relaciones simples, la respuesta podía encontrarse en modelos lineales de bajo número de variables. Sin embargo, en los mercados actuales, la información es abundante y diversa (fundamentales de empresas, datos macroeconómicos, noticias, redes sociales, transacciones de alta frecuencia). En ese contexto, la forma funcional que liga la información con los rendimientos esperados es compleja y difícil de anticipar.

Aquí es donde la IA juega un papel central, la IA ofrece herramientas flexibles capaces de procesar grandes volúmenes de información heterogénea, identificar patrones no lineales y seleccionar automáticamente las variables relevantes. Al mismo tiempo, introduce técnicas de regularización que evitan el sobreajuste, garantizando que las predicciones no se limiten a "memorizar" los datos históri-

cos, sino que capturen relaciones con capacidad de generalización. En consecuencia, la IA no sustituye la lógica de la ecuación anterior, sino que la potencia: ayuda a aproximar mejor las expectativas, $E_t[.]$, y con ello, a estimar de manera más precisa las primas de riesgo y a mejorar la asignación de carteras.

2. Grandes cantidades de información y modelos ambiguos

Desafortunadamente, aplicar la ecuación fundamental que se presentó en la sección anterior no es tan sencillo. La dificultad no está en aceptar que los precios reflejan expectativas, sino en cómo se forman esas expectativas a partir de la información disponible.

En los mercados financieros actuales, los inversores se enfrentan a miles de posibles predictores: ratios contables, datos macroeconómicos, indicadores técnicos, curvas de tipos, *ratings* de crédito, noticias, discursos de bancos centrales, señales de redes sociales, imágenes de satélites que muestran la actividad en puertos o incluso patrones de consumo detectados por transacciones digitales. Este exceso de señales plantea dos problemas clásicos:

- *Demasiada información*: cuando el número de predictores crece de forma explosiva, no resulta evidente cuáles son realmente útiles y cuáles solo añaden ruido.
- *Modelos ambiguos*: aun si tuviéramos claro qué señales usar, desconocemos la forma exacta en que deben combinarse con los precios. ¿Es una relación lineal? ¿Existen efectos no lineales, umbrales, interacciones entre variables?

Kelly y Xiu (2023) destacan que esta es la tensión central de la econometría financiera moderna. La información es abundante, pero la forma funcional es incierta. Dicho de otro modo, sabemos mucho, pero no sabemos exactamente cómo organizar ese conocimiento en un modelo.

Aquí es donde entra en juego la IA. Las técnicas de *ML* y *DL* están diseñadas precisamente para trabajar en entornos con:

- *Alta dimensionalidad*: muchos predictores que pueden ser redundantes o estar correlacionados.
- *Relaciones no lineales*: patrones que no pueden capturarse con una simple regresión lineal.
- *Necesidad de regularización*: evitar que el modelo “memorice” los datos históricos y pierda capacidad de generalización.

Un ejemplo intuitivo: supongamos que queremos predecir el retorno de una acción usando 500 variables posibles. Si usamos una regresión tradicional, el modelo podría colapsar: habría más parámetros que observaciones. En cambio, técnicas relativamente sencillas de *ML*, como *lasso* o un modelo de *random forest*, pueden seleccionar las variables más informativas y descartar las irrelevantes, generando predicciones más estables.

Lo interesante es que, incluso con señales débiles, cuando se agregan de forma regularizada pueden mejorar la predicción. No porque cada señal sea muy potente por sí sola, sino porque la combinación de muchas señales pequeñas puede contener información agregada valiosa. En finanzas no buscamos una “bola de cristal” que lo acierte todo, sino métodos que, al organizar grandes cantidades de información de manera robusta, consigan una pequeña mejora sistemática en la predicción.

En el caso de los reguladores, no importa tanto si un modelo usa dos o doscientas variables, lo relevante es que exista un proceso transparente y regularizado que evite el sobreajuste. En un entorno donde los modelos son necesariamente ambiguos, la gobernanza de datos y modelos pasa a ser tan importante como las propias predicciones.

En este punto, surge una pregunta natural: si los mercados generan enormes cantidades de información y la relación entre esas variables y los precios es incierta, ¿cómo podemos organizar todo ese conocimiento de una manera que sea a la vez interpretable y útil para la predicción? La respuesta tradicional de las finanzas ha sido reducir la complejidad mediante modelos factoriales, que permiten resumir los predictores relevantes en un conjunto manejable de dimensiones comunes. Esta

aproximación lineal marcó un gran avance, aunque, como veremos enseguida, también impone limitaciones que justifican abrir la puerta a relaciones más flexibles y, en última instancia, al uso de la IA.

III. DE PRECIOS A RETORNOS: INTRODUCCIÓN A LOS MODELOS FACTORIALES

Comencemos recuperando la ecuación fundamental de precios:

$$p_t = E_t [m_{t+1} + x_{t+1}],$$

Esta formulación captura bien la idea de que los precios de hoy reflejan predicciones sobre pagos futuros ponderados por las preferencias de los inversores. Para fines prácticos en la industria y también en investigación académica, suele resultar más conveniente reescribir la ecuación en términos de retornos esperados. Si dividimos ambos lados por el precio actual llegamos a la representación clásica:

$$1 = E_t [m_{t+1} + R_{t+1}], \quad [2]$$

donde

$R_{t+1} = \frac{x_{t+1}}{p_t}$, es el rendimiento bruto del activo. Esta condición de no arbitraje significa que, en equilibrio, el retorno esperado de un activo debe compensar su covarianza con el factor de descuento estocástico.

El siguiente paso, clave en el desarrollo empírico, es aproximar el factor de descuento con un conjunto reducido de factores. En la tradición iniciada por Ross con la *Arbitrage Pricing Theory* (Ross, 1976) y extendida por Fama y French (1993), se parte de la idea de que un pequeño número de factores puede resumir los riesgos sistemáticos relevantes para explicar los retornos esperados. Así se llega a una representación lineal:

$$E_t [R_{i,t+1}] = \beta_i' \lambda, \quad [3]$$

donde $E_t [R_{i,t+1}]$ es el retorno esperado del activo i , β_i' son las sensibilidades del activo a los factores y λ son las primas de riesgo asociadas. Esta representación lineal ha sido muy influyente porque permite reducir un universo de miles de activos a unas pocas dimensiones de riesgo común, facilitando tanto la construcción

de carteras como la interpretación económica de los retornos.

El problema, como señalan Kelly y Xiu (2023), es que esta linealidad puede ser demasiado restrictiva. No hay garantía de que los retornos se relacionen siempre de manera estrictamente lineal con los factores, ni de que las primas de riesgo sean constantes en el tiempo. Las interacciones entre factores, los efectos no lineales o los cambios de régimen pueden ser decisivos. En otras palabras, los retornos pueden depender de los factores a través de una función desconocida y posiblemente no lineal:

$$E_t [R_{i,t+1}] = f(\beta_i), \quad [4]$$

donde $f(\cdot)$ es una función flexible que no necesitamos restringir a una forma concreta. Aquí es donde la IA ofrece su mayor valor añadido. Técnicas como redes neuronales, árboles de decisión o métodos de *boosting* permiten aproximar funciones no lineales sin imponer un molde rígido. Lo esencial es que, gracias a la regularización, podemos explorar un espacio mucho más rico de posibles relaciones sin caer en el sobreajuste.

La intuición es clara: si los precios son predicciones y los retornos dependen de múltiples factores, la IA puede ayudarnos a descubrir relaciones complejas entre factores y retornos esperados. Esto abre la puerta a estrategias de inversión más informadas y robustas, en las que los factores sirven como insumos con sentido económico y el ML aporta la flexibilidad necesaria para captar patrones que los modelos lineales tradicionales podrían pasar por alto.

IV. CASO PRÁCTICO: ESTRATEGIA DE INVERSIÓN ADAPTADA, LA INDUSTRIA DE GESTIÓN DE ACTIVOS (LONG-ONLY)

1. Datos y universo de estudio

El análisis empírico que presentamos en este artículo se apoya en datos de FactSet, utilizando como universo las empresas que históricamente han compuesto el índice S&P 500. En concreto, el conjunto de datos proviene de un universo que incluye los componentes del índice extraídos del

CUADRO N.º 1

EMPRESAS DE MAYOR CAPITALIZACIÓN BURSÁTIL DEL UNIVERSO UTILIZADO EN LA APLICACIÓN EMPÍRICA DEL ARTÍCULO, ORDENADAS DE MAYOR A MENOR. "MKTCAP" (CAPITALIZACIÓN BURSÁTIL)

En millones de USD

| TICKER | NOMBRE | ÚLTIMO PRECIO | CAPITALIZACIÓN DE MERCADO (MILLONES USD) | SECTOR | CÓDIGO DE INDUSTRIA | % CAMBIO 4 SEMANAS |
|-----------|--------------------|---------------|--|-------------|---------------------|--------------------|
| NVDA:USA | NVIDIA | 170.76 | 4.157.494 | TECH | SEMIMFR | -6,21 |
| MSFT:USA | Microsoft | 498.41 | 3.705.180 | TECH | SOFTW | -4,48 |
| AAPL:USA | Apple | 234.35 | 3.481.673 | TECH | COMMEQUIP | 3,16 |
| GOOGL:USA | Alphabet | 239.63 | 2.900.482 | TECH | INTERNET | 19,22 |
| AMZN:USA | Amazon | 238.24 | 2.539.638 | NONCYCLICAL | GENMERCH | 7,65 |
| META:USA | Meta Platforms | 765.7 | 1.926.501 | TECH | INTERNET | -0,02 |
| AVGO:USA | Broadcom | 336.67 | 1.583.359 | TECH | SEMIMFR | 10,78 |
| TSLA:USA | Tesla | 346.97 | 1.118.631 | CYCLICALS | VEHICLE | 2,34 |
| BRK.B:USA | Berkshire Hathaway | 492.72 | 1.062.962 | FINANCIAL | INSURANCE | 6,02 |
| JPM:USA | JPMorgan Chase | 297.85 | 819.014 | FINANCIAL | BANKSUSA | 2,86 por 100 |

Nota: Fecha, 10 de septiembre de 2025.

Fuente: Elaboración propia a partir de datos de FactSet.

CUADRO N.º 2

EMPRESAS DE MENOR CAPITALIZACIÓN BURSÁTIL DEL UNIVERSO UTILIZADO EN LA APLICACIÓN EMPÍRICA DEL ARTÍCULO, ORDENADAS DE MENOR A MAYOR. "MKTCAP" (CAPITALIZACIÓN BURSÁTIL)

En millones de USD

| TICKER | NOMBRE | ÚLTIMO PRECIO | CAPITALIZACIÓN DE MERCADO (MILLONES USD) | SECTOR | CÓDIGO DE INDUSTRIA | % CAMBIO 4 SEMANAS |
|----------|----------------------------|---------------|--|--------------|---------------------|--------------------|
| ENPH:USA | Enphase Energy | 37,94 | 4.960 | TECH | COMPONENTS | 16,49 |
| CZR:USA | Caesars Entertainment | 25,39 | 5.281 | CONSUMERSVCE | HOSPITALITY | 5,79 |
| MKTX:USA | MarketAxess Holdings | 188,17 | 7.041 | FINANCIAL | INVESTSVCE | -0,58 |
| CRL:USA | Charles River Laboratories | 157,38 | 7.745 | HEALTHCARE | HCAREOUTSOURCE | 5,67 |
| EMN:USA | Eastman Chemical | 67,6 | 7.759 | MATERIALS | CHEMSPECIAL | 10,69 |
| APA:USA | APA | 21,66 | 7.771 | ENERGY | FOSSILFUEL | 8,73 |
| LW:USA | Lamb Weston Holdings | 57,12 | 7.953 | NONCYCLICAL | FOODDRINKMFR | 10,59 |
| HSIC:USA | Henry Schein | 67,65 | 8.246 | HEALTHCARE | HCARESUPPORT | 2,34 |
| LKQ:USA | LKQ | 32,31 | 8.313 | CYCLICALS | AUTORETAIL | 10,50 |
| TECH:USA | Bio-Techne | 54,2 | 8.399 | HEALTHCARE | HCAREMISC | 8,08 |

Nota: Fecha, 10 de septiembre de 2025.

Fuente: Elaboración propia a partir de datos de FactSet.

ETF IVV, es decir, las 500 compañías de gran capitalización seleccionadas por Standard & Poor's en función de sus criterios de liquidez y representatividad sectorial. Es importante aclarar en este punto que, en la selección del universo, no se incurre en ningún sesgo de supervivencia ni en el uso de información no disponible en el instante de tiempo de confección del universo en el pasado.

La elección de este universo no es casual, ni tampoco la más favorable para quien busca generar rentabilidades superiores al mercado. El S&P 500 está formado por empresas con un altísimo grado de seguimiento por parte de analistas, fondos institucionales y medios de comunicación, lo que implica que sus precios deberían ser más eficientes en el sentido de que reflejan con rapidez y precisión la información disponible. En otras palabras, en este universo existen menos oportunidades de generar "alfa", entendido como rentabilidades ajustadas al riesgo superiores a las del propio mercado.

Precisamente por ello, el hecho de elegir el S&P 500 plantea una prueba exigente para cualquier metodología de predicción y construcción de carteras. Si la IA es capaz de identificar patrones útiles incluso en un conjunto de activos tan competido y analizado, donde la competencia por la información es máxima, entonces sus posibilidades de aportar valor en universos menos líquidos o con menor cobertura analítica serían aún mayores.

Otra ventaja práctica de este universo es que elimina preocupaciones habituales en entornos de menor liquidez. En el S&P 500 los problemas de ejecución, deslizamiento (*slippage*) o impacto de mercado son mínimos, lo que permite aislar mejor la contribución de los modelos a la generación de rentabilidades sin introducir sesgos operativos. Además, trabajar con compañías de gran capitalización facilita la replicabilidad de los resultados y su relevancia para muchas gestoras, dado que se trata del conjunto de referencia para la mayoría de las carteras institucionales globales.

En resumen, usar los datos del S&P 500 a través de FactSet nos sitúa en un terreno donde vencer al índice de referencia es especialmente difícil, lo que convierte al experimento en una prueba de con-

cepto exigente. Si la IA puede añadir valor en este contexto, es razonable pensar que también lo hará en otros universos de activos donde la eficiencia de precios sea menor y las oportunidades de predicción más abundantes.

Para finalizar la sección —y, a modo puramente ilustrativo—, me gustaría presentar algunas de las empresas que conformarían actualmente el universo. A fecha 10 de septiembre de 2025, la composición del universo elegido en el artículo refleja de manera clara la concentración del mercado en las grandes compañías tecnológicas. En el cuadro n.º 1 se muestran las diez mayores empresas por capitalización bursátil en nuestro universo, con NVIDIA, Microsoft y Apple a la cabeza, todas ellas con valoraciones que superan los tres billones de dólares. Este grupo se completa con gigantes como Alphabet, Amazon y Meta, a los que se suman Broadcom y Tesla en el sector tecnológico e industrial, y dos pesos pesados no tecnológicos: Berkshire Hathaway y JPMorgan.

El contraste con el cuadro n.º 2, donde se recogen las compañías de menor tamaño dentro del índice, resulta ilustrativo. Allí encontramos firmas con capitalizaciones inferiores a los diez mil millones de dólares, pertenecientes a sectores más fragmentados como energía, consumo, salud y materiales. Esta heterogeneidad, desde líderes globales con enorme liquidez hasta empresas medianas con seguimiento más limitado, muestra la riqueza del universo S&P 500 y refuerza la idea de que identificar oportunidades de inversión en este conjunto exige modelos capaces de discriminar señales en entornos muy distintos de cobertura y eficiencia informativa.

2. Conjunto de factores utilizados para la predicción

Para la predicción de retornos utilizamos un conjunto de 65 factores derivados de la base de datos de FactSet. Estos factores se han agrupado en cuatro grandes familias, siguiendo la lógica de los modelos factoriales discutidos previamente y extendiéndola a un entorno de mayor dimensionalidad. El uso de un número elevado de predictores se ha convertido en práctica habitual en la literatura reciente sobre valoración de activos (*asset pricing*),

donde los llamados “factor zoo” han motivado la aplicación de métodos de ML capaces de seleccionar y ponderar decenas o incluso centenares de señales simultáneamente (Harvey *et al.*, 2016; Gu *et al.*, 2020; Bryzgalova *et al.*, 2023).

Value & Growth factors (factores basados en valor y crecimiento): incluyen medidas fundamentales de beneficios, ventas y crecimiento. Por ejemplo, siguiendo la notación de FactSet para su posible reproducibilidad:

- “EPSExcIXorGr%PYQ”: crecimiento trimestral del beneficio por acción (excluyendo partidas extraordinarias) respecto al mismo trimestre del año anterior.
- “EPSExcIXorGr%TTM”: crecimiento del beneficio por acción acumulado en los últimos doce meses frente al mismo período del año anterior.
- “EPSExcIXorGr5Y”: crecimiento compuesto anual del beneficio por acción en los últimos cinco años.

Price factors (factores basados en precios). Capturan la dinámica histórica de precios y volatilidad. Por ejemplo:

- *Price Volatility, 12 months*: desviación estándar de los rendimientos diarios de la acción durante los últimos doce meses.
- *Price Volatility, 3 years*: volatilidad calculada sobre un horizonte móvil de tres años.
- *Price Volatility, 5 years*: medida equivalente en una ventana de cinco años, que capta estabilidad relativa a largo plazo.

Vol-Risk factors (factores basados en volatilidad y riesgo). Reflejan la sensibilidad de la acción al mercado en diferentes horizontes. Por ejemplo:

- *1-year Beta*: beta de la acción calculada frente al índice de referencia en los últimos doce meses.
- *Beta3Y*: sensibilidad media al mercado en un horizonte de tres años.

- *Beta5Y*: versión ampliada a cinco años, que recoge estabilidad o cambios en el perfil de riesgo sistemático.

Sentiment factors (factores basados en sentimiento de analistas). Incorporan percepciones de mercado y señales cualitativas. Por ejemplo:

- *Changes in recommendation* (cambios en recomendación de analistas): recoge la dirección y magnitud de los cambios recientes en las recomendaciones de analistas sobre la acción (mejoras o rebajas de *rating*).

El propósito de esta clasificación es mostrar cómo los factores funcionan como bloques de construcción de modelos predictivos. Mientras los enfoques tradicionales se centran en un número reducido de factores lineales, aquí ampliamos el abanico a decenas de señales que la IA puede procesar de manera conjunta, explorando relaciones no lineales y combinaciones complejas.

Para que los valores de estos factores sean comparables entre sí y puedan usarse como insumos en modelos predictivos, se aplica un proceso de normalización en forma de *ranking*. Concretamente, los valores se ordenan en orden descendente y se les asigna un valor que va desde 1 hasta -1, de modo que los valores más altos reciben puntuaciones cercanas a 1 y los más bajos puntuaciones cercanas a -1. El paso entre percentiles se calcula excluyendo los valores perdidos (NAs), a los que se asigna el valor cero.

Este procedimiento asegura que todos los factores, independientemente de sus unidades originales (por ejemplo, porcentajes de crecimiento, desviaciones estándar o cambios en recomendaciones), se expresen en una escala común. De esta forma, es posible compararlos, combinarlos y utilizarlos como *inputs* de modelos de aprendizaje automático sin que las diferencias de escala distorsionen el entrenamiento o el peso relativo de cada señal.

Por razones de espacio no se incluye en el artículo la definición detallada de los 65 factores utilizados, aunque el listado completo se encuentra disponible contactando directamente con el autor.

3. Factores y su análisis univariante por deciles

Antes de presentar el modelo de *machine learning*, ML, que combina los 65 factores simultáneamente, resulta ilustrativo mostrar cómo se trabaja tradicionalmente con un factor de forma aislada. Este enfoque, muy extendido en la literatura académica, consiste en ordenar los activos cada mes según el valor de un factor, formar carteras por deciles y analizar su rendimiento a lo largo del tiempo.

En concreto, se procede así:

- *Ranking mensual*. A final de cada mes, ordenamos las 500 compañías del universo en función del valor de un factor, previamente normalizado en un rango comparable.
- *Formación de carteras*. Dividimos el universo en 10 grupos (deciles). El decil 1 contiene las compañías con valores más bajos del factor, y el decil 10 las que presentan valores más altos.
- *Rebalanceo dinámico*. Al comienzo del mes siguiente, se forman carteras igualmente ponderadas dentro de cada decil y se mantienen durante un mes. En el mes siguiente se repite el proceso, actualizando *rankings* y carteras.
- *Medición de resultados*. Calculamos la rentabilidad media de cada decil, cartera, en el periodo enero 2005–septiembre 2025. El interés está en observar si los deciles superiores (10) superan sistemáticamente a los inferiores (1).

En esta primera ilustración utilizamos dos factores representativos de familias distintas:

- *Factor de valor*: “OpIncBDeprTTM / EV”. Esta ratio es análoga a un *ebitda* (acrónimo de los términos en inglés *earnings before interest taxes depreciation and amortization*) sobre valor de empresa. Mide la rentabilidad operativa de la compañía en relación con su valoración total. La intuición económica es sencilla: las empresas que generan más beneficios operativos por unidad de valor de mercado deberían ofrecer mejores retornos futuros, al estar “infravaloradas” en términos relativos. Este tipo de ratios de valoración se han utilizado

de manera sistemática en la literatura empírica de *asset pricing* como predictores de la sección cruzada (*cross-section*) de rentabilidades, mostrando que medidas basadas en beneficios operativos o flujos de caja, ajustadas por el valor de mercado, contienen información sobre primas de riesgo y oportunidades de inversión (Lakonishok et al., 1994; Fama y French, 1998; Hou et al., 2012).

- *Factor de sentimiento*: “EPS Estimate Revision CY”. Este indicador mide el cambio porcentual en las estimaciones de los analistas de los beneficios por acción (EPS) para el año fiscal corriente, comparando las estimaciones actuales con las de cuatro semanas atrás. La lógica es que las revisiones al alza en las expectativas de analistas reflejan nueva información positiva y tienden a anticipar un mejor comportamiento bursátil. La literatura académica ha documentado consistentemente que las revisiones de beneficios constituyen uno de los predictores más sólidos en el *cross-section* de retornos, al recoger la incorporación gradual de información privada y pública en los precios (Givoly y Lakonishok, 1979; Abarbanell y Bernard, 1992; Chan et al., 1996).

Para ilustrar de manera concreta cómo se construyen estos *rankings*, en el cuadro n.º 3 mostramos las diez compañías con mayor puntuación en el factor “OpIncBDeprTTM/EV” a fecha 10 de septiembre de 2025, mientras que el cuadro n.º 4 recoge las diez primeras según el factor “EPS Estimate Revision CY”. Para facilitar la interpretación de los cuadros, el *ranking* se presenta en la escala [0, 100] en lugar de en la escala original [-1, 1]. No obstante, el modelo de aprendizaje automático utiliza la escala [-1, 1], tal como se explicó anteriormente. Estos cuadros permiten visualizar qué tipo de empresas se sitúan en la parte más alta de la distribución, donde, según la intuición teórica, se concentran las oportunidades de generación de rentabilidad.

Una vez visto el *ranking* puntual, el siguiente paso es evaluar la validez de estas señales en el tiempo. Para ello, cada mes desde enero de 2005 hasta septiembre de 2025 ordenamos las compañías según el *ranking* determinado por cada factor, formamos 10 carteras igualmente ponderadas (de-

CUADRO N.º 3

RANKING (RANK) DE EMPRESAS DENTRO DEL UNIVERSO ORDENADAS DE MAYOR A MENOR POR EL FACTOR SENTIMIENTO: "EPS ESTIMATE REVISION CY" (CAPITALIZACIÓN BURSÁTIL)

En millones de USD

| TICKER | SECTOR | CAPITALIZACIÓN DE MERCADO (MILLONES USD) | RANK % |
|---------|------------------------|--|--------|
| FRT:USA | Finance | 8.655 | 99,8 |
| WBD:USA | Consumer Services | 30.344 | 99,6 |
| ALB:USA | Non-Energy Materials | 8.540 | 99,4 |
| CPT:USA | Finance | 11.928 | 99,2 |
| DG:USA | Consumer Non-Cyclicals | 23.012 | 99 |
| EQR:USA | Finance | 25.314 | 98,8 |
| ESS:USA | Finance | 17.316 | 98,6 |
| APA:USA | Energy | 7.771 | 98,4 |
| KIM:USA | Finance | 15.142 | 98,2 |
| AVB:USA | Finance | 27.799 | 98 |

Nota: Fecha, 10 de septiembre de 2025.

Fuente: Elaboración propia a partir de datos de FactSet.

CUADRO N.º 4

RANKING (RANK) DE EMPRESAS DENTRO DEL UNIVERSO ORDENADAS DE MAYOR A MENOR POR EL FACTOR DE VALOR: "OPINCBDEPRTTM / EV". (CAPITALIZACIÓN BURSÁTIL)

En millones de USD

| TICKER | SECTOR | CAPITALIZACIÓN DE MERCADO (MILLONES USD) | RANK % |
|----------|--------------------|--|--------|
| BK:USA | Finance | 73.888 | 99,8 |
| APA:USA | Energy | 7.771 | 99,6 |
| DVN:USA | Energy | 21.821 | 99,4 |
| MOH:USA | Healthcare | 9.803 | 99,2 |
| SYF:USA | Finance | 28.294 | 99 |
| EOG:USA | Energy | 64.050 | 98,8 |
| HAL:USA | Energy | 18.553 | 98,6 |
| GM:USA | Consumer Cyclicals | 55.516 | 98,4 |
| CTRA:USA | Energy | 18.198 | 98,2 |
| OXY:USA | Energy | 44.437 | 98 |

Nota: Fecha, 10 de septiembre de 2025.

Fuente: Elaboración propia a partir de datos de FactSet.

ciles), y mantenemos cada cartera durante un mes, rebalanceando en el siguiente. En los gráficos 2 y 3 se muestran las rentabilidades acumuladas de cada decil. En ambos factores, el decil 10, empresas con los valores más altos del factor, ofrece retornos claramente superiores al decil 1, lo que confirma que incluso analizados de manera univariante, estos factores pueden contener información predictiva sobre los rendimientos futuros.

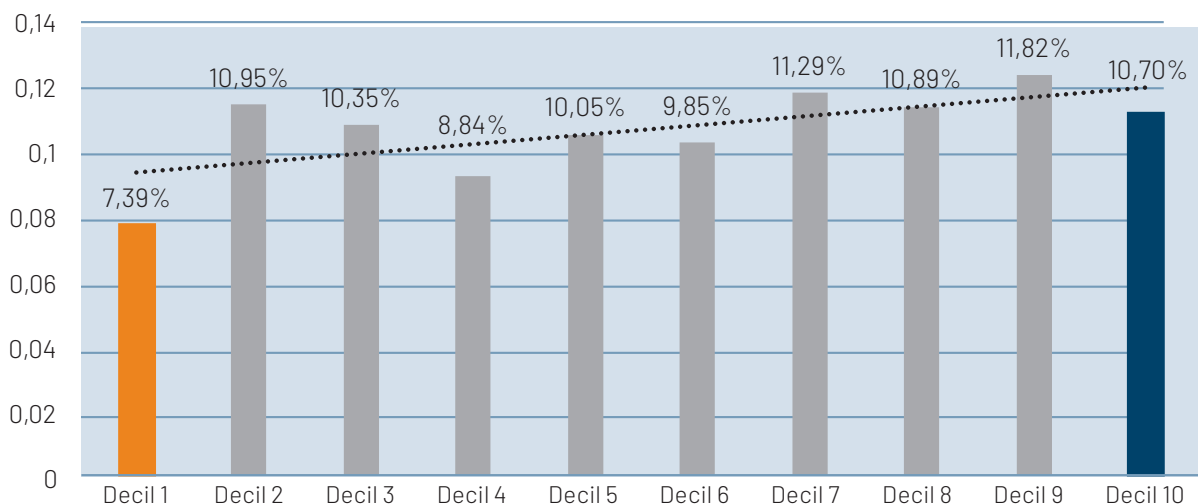
En el gráfico 2 se presentan las rentabilidades anuales promedio de las carteras formadas según el factor "EPS Estimate Revision CY". Se aprecia una diferencia clara entre el primer decil, con una rentabilidad anual del 7,39 por 100, y el décimo decil, con un 10,70 por 100. Aunque el diferencial no es muy amplio, la señal es coherente con la intuición: las compañías con revisiones más positivas en sus estimaciones de beneficios tienden a generar mejores resultados bursátiles. Cabe destacar que el período considerado, 2005-2025, ha sido es-

pecialmente favorable para el S&P 500 en su conjunto, lo que explica que incluso los deciles menos favorecidos presenten rentabilidades elevadas. En línea con este hecho empírico, en la figura también se observa la rentabilidad anualizada del *benchmark* (S&P 500) que es del 10,50 por 100.

El gráfico 3 muestra los resultados equivalentes para el factor "OpInCBDeprTTM / EV", y en este caso la diferencia entre extremos es aún más marcada: mientras el primer decil obtiene un 6,59 por 100 de rentabilidad anual, el décimo decil alcanza un 11,88 por 100. Además, como se desprende del gráfico, la rentabilidad anualizada del décimo decil es superior a la del *benchmark* también, 10,50 por 100. Este patrón es consistente con la idea de que las compañías con mayor capacidad de generar beneficios operativos en relación con su valor de mercado tienden a ofrecer un desempeño superior. La comparación entre ambos gráficos refuerza la noción de que, incluso

GRÁFICO 2

RETORNOS ANUALES POR DECILES USANDO EL FACTOR "EPS ESTIMATE REVISION CY"

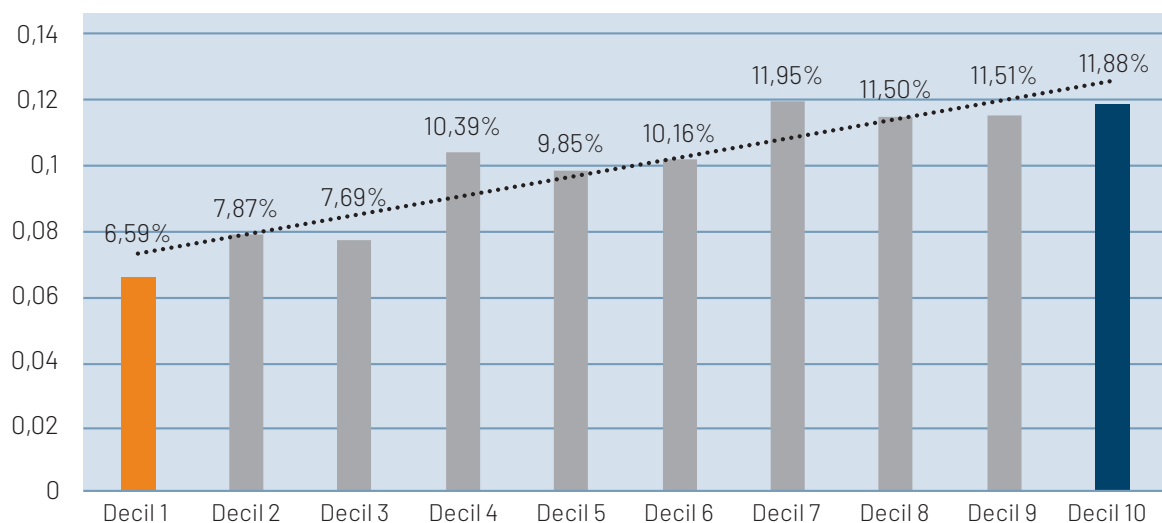


Nota: Rebalanceo mensual. Desde 2005 hasta 2025.

Fuente: Elaboración propia a partir de datos de FactSet.

GRÁFICO 3

RETORNOS ANUALES POR DECILES USANDO EL FACTOR "OPINCBDEPRTM / EV"



Nota: Desde 2005 hasta 2025.

Fuente: Elaboración propia a partir de datos de FactSet.

analizados de forma univariante y en un mercado altamente seguido como el S&P 500, los factores aportan señales útiles y persistentes sobre los retornos futuros.

Este análisis preliminar se presenta sin costes de transacción ni comisiones. La exclusión es deliberada, ya que el objetivo aquí es ilustrar la lógica de los factores en su forma más sencilla. En el

ejercicio posterior, donde se combinan múltiples factores mediante técnicas de IA, incorporaremos dichos costes para ofrecer una evaluación más realista.

4. Modelo de IA utilizado: impulso por gradiente (*gradient boosting*)

En la parte teórica hemos introducido la ecuación:

$$E_t [R_{i,t+1}] = f(\beta_i),$$

que resume la idea de que los rendimientos esperados de un activo pueden expresarse como una función de un conjunto de factores. Siguiendo este marco, en el análisis empírico utilizamos un vector β_i compuesto por los 65 factores previamente descritos, que abarcan dimensiones de valor, crecimiento, precio, riesgo y sentimiento. Cada mes estos factores se normalizan mediante un *ranking* en el intervalo $[-1,1]$, excluyendo valores no válidos y asignando a estos últimos un 0, con el fin de asegurar comparabilidad y consistencia entre señales heterogéneas. La función, $f(\cdot)$, se aproxima a través de un modelo de impulso por gradiente (*gradient boosting*) sobre árboles de decisión, concretamente el algoritmo “*LightGBM*”, ampliamente utilizado en predicción financiera por su capacidad de manejar grandes volúmenes de información, capturar interacciones y no linealidades y evitar sobreajuste mediante regularización (Ke et al., 2017). Este diseño permite captar patrones débiles pero persistentes en entornos de señal-ruido desfavorables como los que caracterizan a los mercados financieros (Kelly y Xiu, 2023).

5. Entrenamiento del modelo de IA

La elección del modelo y de sus parámetros se llevó a cabo en dos etapas. En una primera fase se empleó un esquema de validación cruzada bloqueada (*k-fold cross validation*) adaptado a series temporales, siguiendo los principios de depuración (*purging*) y embargos (embargo) descritos por López de Prado (2018). El objetivo en este punto no es entrenar el modelo definitivo, sino identificar un algoritmo y un conjunto de parámetros razonables para trabajar con los 65 factores definidos previa-

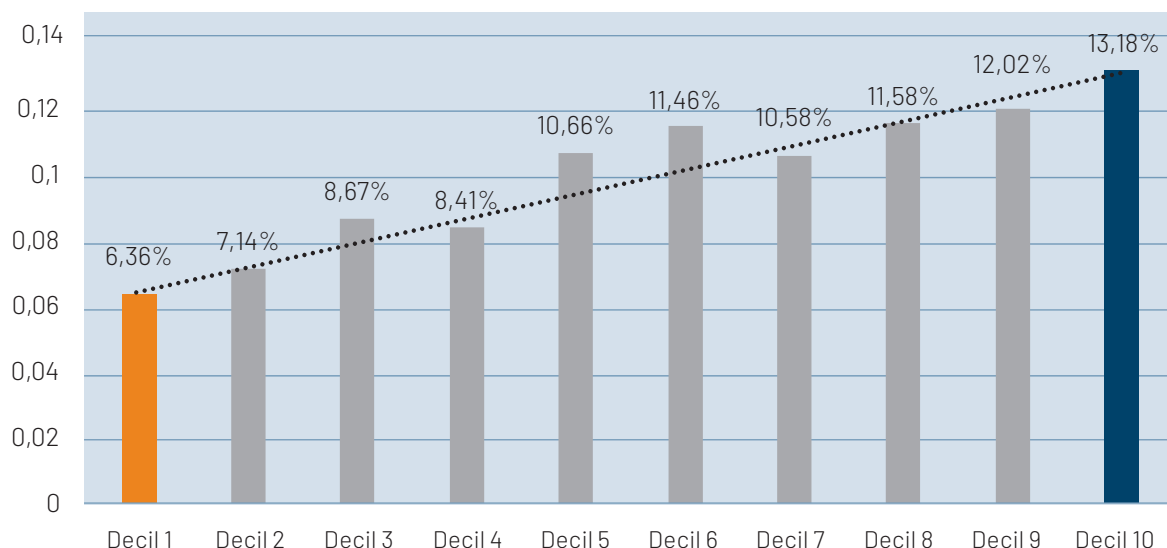
mente. Esta estrategia asegura que la selección no dependa de información futura y que el rendimiento del modelo sea estable a lo largo de distintos subperíodos históricos. En la práctica, el procedimiento consiste en dividir la muestra en varias ventanas temporales consecutivas, de modo que cada bloque de validación se evalúa únicamente sobre períodos no utilizados en el entrenamiento previo. Esta separación temporal evita fugas de información y permite comprobar la robustez del modelo en distintos contextos de mercado, garantizando que las señales extraídas no son fruto de un ajuste espurio a un único período histórico.

El procedimiento empírico dentro de esta primera etapa consiste en dividir la muestra en varias ventanas temporales recurrentes, de manera que el modelo se entrena siempre con información disponible hasta un momento t y se evalúa en un bloque posterior que queda reservado como período de validación. Así, en cada pliegue se generan predicciones $\hat{E}_t [R_{i,t+1}]$ un mes hacia adelante únicamente para el intervalo de validación correspondiente, garantizando que no exista contaminación de información futura. Este diseño permite comprobar de forma repetida cómo se comporta el algoritmo en distintos subperíodos del S&P 500, incluyendo fases de expansión, crisis y recuperación, y asegura que los resultados no dependan de un período específico del ciclo de mercado. En esencia, el esquema alterna bloques de entrenamiento y validación de forma sistemática, de modo que cada evaluación se realiza siempre sobre observaciones no utilizadas previamente en la calibración del modelo.

El propósito de este ejercicio no es obtener un modelo definitivo, sino contar con una base empírica sólida para elegir un algoritmo y un conjunto de parámetros razonables que funcionen de manera estable a lo largo de distintos períodos históricos. La validación cruzada bloqueada permite observar cómo varía el rendimiento del modelo en fases muy diferentes del mercado, incluyendo expansiones, recesiones y episodios de alta volatilidad, y descartar configuraciones que solo resulten adecuadas en intervalos concretos. De esta forma, el procedimiento garantiza que la selección del modelo no dependa de un episodio particular ni de informa-

GRÁFICO 4

RETORNOS ANUALES POR DECILES USANDO EL MODELO DE ML EN EL PERÍODO DESDE 2005 HASTA 2025



Fuente: Elaboración propia a partir de datos de FactSet.

ción futura, sentando las bases para el segundo paso, en el que se reentrenará el algoritmo con un corte temporal fijo para generar predicciones estrictamente fuera de muestra.

El gráfico 4 muestra la rentabilidad anualizada media obtenida en los distintos períodos de validación cruzada cuando las compañías se ordenan en deciles según las predicciones del modelo. El patrón es claro: el decil 1, que agrupa a las empresas con menores retornos esperados, alcanza, únicamente, un 6,36 por 100, mientras que el decil 10 llega al 13,18 por 100. La pendiente ascendente entre extremos confirma que las señales generadas contienen información predictiva consistente en todos los bloques temporales considerados. El promedio del universo, con un 10,13 por 100, y la referencia del índice, con un 10,01 por 100, aparecen representados como líneas horizontales y sirven para contextualizar los resultados: aunque la mayoría de los deciles se sitúan en torno a la rentabilidad del mercado, los extremos, en particular el decil superior, muestran diferencias significativas,

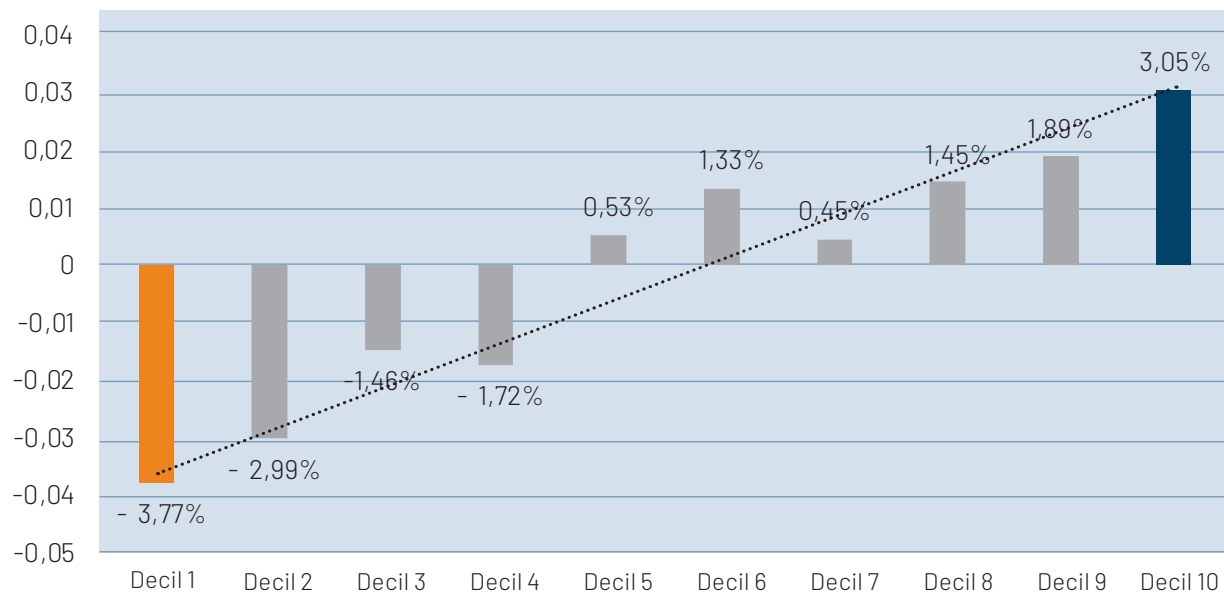
lo que refuerza la utilidad del modelo incluso en esta primera etapa exploratoria.

El gráfico 5 presenta la rentabilidad de cada decil en exceso frente al *benchmark*. Se observa que los deciles intermedios se sitúan próximos a cero, mientras que el decil 1 muestra una rentabilidad relativa claramente negativa (-3,77 por 100) y el decil 10 un rendimiento superior al índice (+3,05 por 100). Este perfil monótono confirma que el modelo es capaz de discriminar de manera consistente entre las compañías con menores y mayores retornos esperados.

El gráfico 6 compara la evolución de las carteras de los deciles extremos frente al *benchmark* durante todos los períodos de validación cruzada. El decil 1 (línea naranja) muestra de forma consistente un rendimiento inferior al índice, mientras que el decil 10 (azul claro) se sitúa de manera persistente por encima de la referencia (línea negra). Aunque las diferencias se amplifican o reducen según el entorno de mercado, el perfil general es estable y

GRÁFICO 5

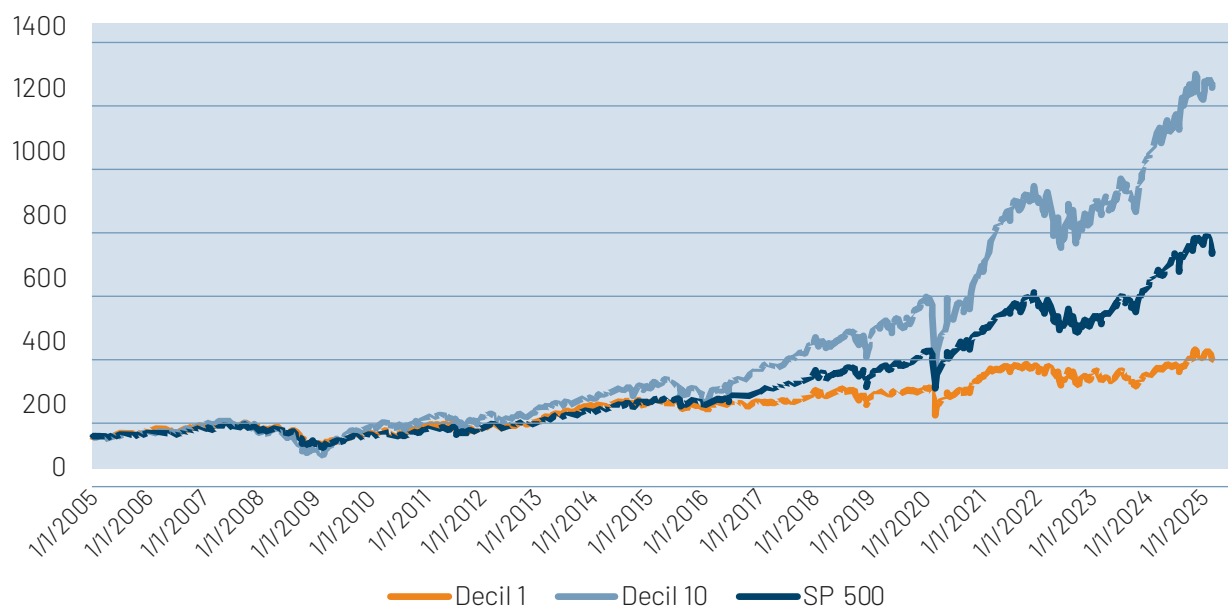
RETORNOS ANUALES EN EXCESO DEL SP 500 POR DECILES USANDO EL MODELO DE ML EN EL PERÍODO DESDE 2005 HASTA 2025



Fuente: Elaboración propia a partir de datos de FactSet.

GRÁFICO 6

RETORNOS ACUMULADOS USANDO EL MODELO DE ML EN EL PERÍODO DESDE 2005 HASTA 2025. EN AZUL CLARO, RENTABILIDAD DEL DECIL 10, EN NARANJA EL DEL DECIL 1 Y EN AZUL OSCURO LA DEL SP 500



Fuente: Elaboración propia a partir de datos de FactSet.

confirma que el modelo discrimina con éxito entre las empresas con menores y mayores retornos esperados, reforzando así los resultados observados en las métricas resumidas por deciles.

La elección del modelo no se ha basado únicamente en criterios de rentabilidad, sino también en métricas estadísticas de predicción como el *RMSE* (*“root mean squared error”*, raíz del error cuadrático Medio) y en indicadores de rentabilidad ajustada al riesgo, incluyendo la ratio de Sharpe y la ratio de información. Aunque por cuestiones de espacio no se reporten estos resultados en detalle, sí han sido tenidos en cuenta junto con la evaluación de estabilidad en distintos subperíodos. De este modo, la selección refleja tanto la capacidad predictiva como la robustez económica del modelo, más allá del simple rendimiento bruto.

En conjunto, este proceso de validación cruzada y de selección basada en criterios múltiples garantiza que el modelo elegido no solo maximice la rentabilidad, sino que también conserve propiedades estadísticas sólidas y un perfil de riesgo ajustado. Con ello, estamos en condiciones de conectar directamente el marco teórico planteado con la propuesta empírica, mostrando cómo los factores se integran en una función $f(\cdot)$ flexible que puede ser utilizada para construir estrategias de inversión aplicables en la práctica.

La aproximación empírica que presentamos para $f(\cdot)$, constituye así una operacionalización directa de la propuesta de Kelly y Xiu (2023). Pasaremos de una formulación teórica en la que los precios son predicciones condicionadas a la información disponible, a un ejercicio aplicado en el que los factores previamente definidos se introducen en un modelo flexible y regularizado, capaz de extraer señal de un conjunto amplio de variables y traducirla en estrategias de inversión implementables. De esta forma, el análisis empírico no sustituye a la teoría, sino que la prolonga en un marco donde la abundancia de información y la ambigüedad de las formas funcionales hacen conveniente recurrir a metodologías de *ML*.

Una vez comprobado que el modelo funciona correctamente y ofrece señales consistentes, se

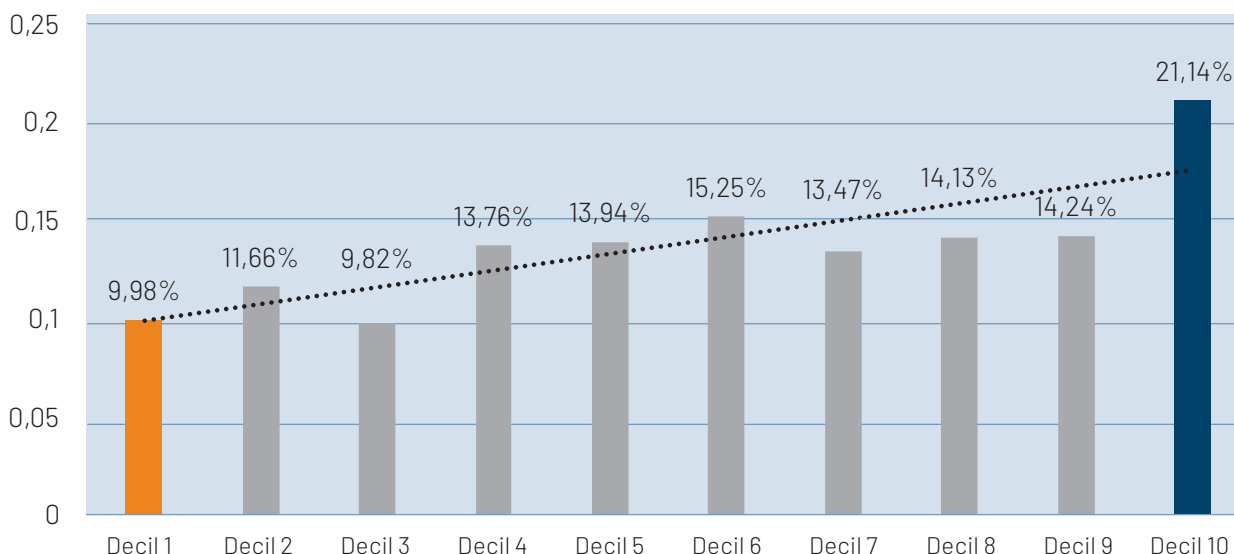
llevó a cabo un segundo entrenamiento empleando únicamente los datos disponibles hasta el 31 de diciembre de 2019. Este modelo entrenado durante esta segunda etapa será el usado fuera de la muestra. Con esta estrategia, todos los resultados que presento a continuación corresponden al período 2020-2025, considerado estrictamente como fuera de muestra (*“out-of-sample”*, *OOS*), lo que refuerza la credibilidad y la validez del ejercicio.

6. Evaluación del modelo fuera de la muestra

La evaluación fuera de la muestra se realiza sobre el período comprendido entre el 11 de septiembre de 2020 y el 11 de septiembre de 2025, lo que asegura que no existe solapamiento alguno con los datos empleados en el entrenamiento del modelo, que se limitó hasta el 31 de diciembre de 2019. Es importante subrayar que la validación cruzada bloqueada utilizada en la etapa previa, aunque abarca toda la muestra histórica, no constituye una violación del principio de análisis fuera de la muestra. Su propósito es exclusivamente la selección de un algoritmo y un conjunto de parámetros robustos, sin que los resultados obtenidos en esa fase se reporten como evidencia predictiva definitiva. Dicho de otro modo, el ejercicio de *k-fold cross validation* sirve para evaluar la estabilidad del modelo en distintos subperíodos históricos, pero la verdadera prueba de generalización se produce en este tramo *OOS*, que no intervino en ningún momento en la calibración ni en la selección del modelo. Adicionalmente, por robustez, incluso restringiendo la primera etapa al 31 de diciembre de 2019, los resultados no varían sustancialmente. De este modo, los resultados que presentamos a continuación ofrecen una medida limpia y creíble de la capacidad predictiva del enfoque en un entorno genuinamente prospectivo.

El gráfico 7 presenta los resultados por deciles durante el período estrictamente fuera de la muestra, del 11 de septiembre de 2020 al 11 de septiembre de 2025. El patrón es claro: el decil 1, que agrupa a las compañías con predicciones más bajas, obtiene una rentabilidad anualizada del 9,98 por 100, mientras que el decil 10 alcanza un 21,14 por 100, muy por encima del *benchmark* de referencia (15,65 por 100). La pendiente ascendente entre extremos confirma que el modelo discrimina con éxito entre

GRÁFICO 7
RETORNOS ANUALES POR DECILES USANDO EL MODELO DE ML FUERA DEL PERÍODO DE ENTRENAMIENTO,
DESDE 9/11/2020 HASTA 09/11/2025



Fuente: Elaboración propia a partir de datos de FactSet.

empresas con mayor y menor retorno esperado, y que esta señal se traduce en diferencias económicas significativas en un entorno prospectivo. Estos resultados refuerzan la validez del enfoque y muestran que, incluso en un mercado tan seguido y competitivo como el S&P 500, el uso de un conjunto amplio de factores en un modelo flexible puede generar información predictiva útil.

7. Estrategia de inversión fuera de la muestra

Antes de presentar las métricas resumidas de nuestra estrategia de inversión conviene precisar el diseño de la estrategia evaluada. Aunque en el análisis preliminar por deciles se han mostrado los resultados para grupos de cincuenta valores (número de acciones en un decil sobre 500 empresas), la estrategia que reportamos en detalle a continuación se implementa de forma más concentrada, seleccionando cada mes las veinte acciones con mayor retorno esperado estimado por el modelo. Esta decisión responde tanto a la práctica habitual de la industria como a la necesidad de maximizar la expresividad de la señal, evitando al mismo tiempo un apalancamiento excesivo en el extremo supe-

rior del *ranking*. Asimismo, todos los resultados se reportan tras incorporar costes de ejecución realistas: una comisión de 0,005 dólares por acción, un *slippage* equivalente al 0,1 por 100 del importe total de cada operación, y un precio de transacción aproximado como el promedio del máximo, mínimo y dos veces el cierre del día siguiente. De este modo, las métricas que se presentan no reflejan una simulación idealizada, sino un escenario cercano a lo que podría obtener un inversor en condiciones de mercado plausibles.

El cuadro n.º 5 resume las principales métricas de desempeño del modelo en el período fuera de la muestra, comparadas con el S&P 500 como referencia. El modelo logra una rentabilidad total del 164,77 por 100, equivalente a una rentabilidad anualizada del 21,50 por 100, claramente superior al 15,96 por 100 del índice. Esta diferencia se traduce en un alfa anualizado de 6,55 por 100, lo que evidencia que las predicciones aportan valor más allá de la mera exposición al mercado. El riesgo asumido, medido por la desviación estándar, es mayor que el del S&P 500 (19,79 por 100 frente a 15,95 por 100), aunque la ratio de Sharpe (1,02) y el de Sortino (1,51)

CUADRO N.º 5
EN LA TABLA SE MUESTRAN DATOS ESTADÍSTICOS
DE LA ESTRATEGIA (“MODELO”) E ÍNDICE S&P 500
FUERA DEL PERIODO DE ENTRENAMIENTO, DESDE
9/11/2020 HASTA 09/11/2025)

En porcentaje

| MÉTRICA | MODELO | S&P 500 |
|------------------------------|--------|---------|
| Retorno total % | 164,77 | 109,68 |
| Retorno anualizado% | 21,5 | 15,96 |
| Caída máxima % | -22,57 | -24,5 |
| Número de muestras mensuales | 59 | 59 |
| Desviación estándar% | 19,79 | 15,95 |
| Ratio de Sharpe | 1,02 | 0,81 |
| Ratio de Sortino | 1,51 | 1,08 |
| Correlación con el benchmark | 0,85 | - |
| R ² | 0,73 | - |
| Beta | 1,06 | - |
| Alfa (anualizado)% | 6,55 | |

Fuente: Elaboración propia a partir de datos de FactSet.

son también más elevados que los del índice (0,81 y 1,08 respectivamente), lo que refleja una mejor relación rentabilidad-riesgo. En términos de caídas máximas, el modelo presenta un *drawdown* del -22,57 por 100, algo menor que el -24,50 por 100 del *benchmark*, lo que muestra una capacidad moderada de protección en entornos adversos. La correlación con el índice (0,85) y una beta ligeramente superior a 1,06 indican que la estrategia mantiene una exposición relevante al mercado, pero con un componente idiosincrático que explica parte de la rentabilidad extra, como sugiere también el R^2 de 0,73. En conjunto, estas métricas confirman que el modelo no solo mejora la rentabilidad absoluta, sino también el perfil ajustado por riesgo en un entorno estrictamente prospectivo.

El contraste entre la estrategia y el índice de referencia en el periodo fuera de la muestra es nítido. La cartera gestionada por el modelo alcanza una rentabilidad anualizada del 21,5 por 100, es decir, 5,5 puntos porcentuales por encima del S&P 500. En términos de capital acumulado, por dar algunas cifras representativas, una inversión inicial de

CUADRO N.º 6
COMPOSICIÓN DE LA CARTERA ML.
FECHA: 11 DE SEPTIEMBRE DE 2025.
En porcentaje

| N.º | TICKER | NOMBRE | PESO |
|-----|--------|---------------------------------|------|
| 1 | CAH | Cardinal Health | 6.8 |
| 2 | EL | The Estée Lauder Companies | 6.7 |
| 3 | SJM | The J. M. Smucker Co. | 6.3 |
| 4 | MDLZ | Mondelez International | 6.2 |
| 5 | ED | Consolidated Edison | 6.2 |
| 6 | HRL | Hormel Foods | 6.1 |
| 7 | MOH | Molina Healthcare | 6.1 |
| 8 | IBM | International Business Machines | 5.9 |
| 9 | KDP | Keurig Dr Pepper | 5.9 |
| 10 | PANW | Palo Alto Networks | 5.6 |
| 11 | ON | ON Semiconductor | 5.6 |
| 12 | CMCSA | Comcast | 4.9 |
| 13 | KMB | Kimberly-Clark | 4.8 |
| 14 | VLO | Valero Energy | 4.5 |
| 15 | CLX | The Clorox Co. | 4.4 |
| 16 | KVUE | Kenvue | 4.2 |
| 17 | BG | Bunge Global SA | 3.6 |
| 18 | HUM | Humana | 2.9 |
| 19 | CNC | Centene | 2.1 |
| 20 | MRNA | Moderna | 0.8 |

Fuente: Elaboración propia a partir de datos de FactSet.

10.000 dólares habría crecido hasta 26.477 dólares con el modelo *ML*, frente a 20.968 dólares en el *benchmark*. El desglose anual corrobora la regularidad del exceso de retorno: destaca 2021, con una ventaja de +26,4 puntos, y 2023, con +12,1 puntos, mientras que, en 2022, un año claramente adverso para los mercados, la estrategia retrocede ligeramente más que el índice (-18,2 por 100 frente a -18,1 por 100).

A modo de cierre, resulta ilustrativo observar la composición concreta de la cartera a fecha 11 de septiembre de 2025, presentada en el cuadro n.º 6. Aunque se trata de un aspecto secundario frente al análisis agregado de rentabilidades y métricas de riesgo, la lista de veinte valores muestra la diversidad sectorial que caracteriza a la estrategia: desde compañías de consumo estable como The J. M. Smucker o Mondelez, hasta tecnológicas consolidadas como

IBM y Palo Alto Networks, pasando por energéticas como “Valero” o sanitarias como Molina Healthcare y Humana. Esta heterogeneidad refleja que el modelo no se concentra en un único sector, sino que selecciona de manera dinámica aquellas empresas que, en cada momento, presentan las señales más favorables dentro del universo del S&P 500. Aunque las posiciones específicas cambian mes a mes en función de los factores, la cartera ejemplifica cómo las predicciones del modelo se traducen en decisiones de inversión tangibles.

V. CONCLUSIÓN Y RIESGOS DEL USO DE LA IA

Los resultados de este artículo muestran que la IA, aplicada sobre un conjunto amplio de factores y con técnicas de aprendizaje automático flexibles, puede generar información predictiva valiosa incluso en un universo tan seguido y competitivo como el S&P 500. La estrategia implementada en el período fuera de la muestra logra un exceso de rentabilidad frente al índice, acompañado de ratios ajustados al riesgo más favorables, lo que refuerza la idea de que los precios, entendidos como predicciones, pueden modelarse de manera más eficaz con herramientas basadas en la IA.

No obstante, el potencial de la IA en los mercados financieros viene acompañado de riesgos que conviene no subestimar. En primer lugar, el carácter de “caja negra” de modelos como los *boosted trees* o las redes neuronales dificulta la interpretabilidad y plantea desafíos para la gestión del riesgo y el cumplimiento regulatorio. En segundo lugar, la homogeneidad en el uso de algoritmos similares puede dar lugar a comportamientos sincronizados que amplifiquen episodios de estrés, incrementando la prociclicidad y el riesgo sistémico. Finalmente, la dependencia creciente de infraestructuras tecnológicas y proveedores externos de datos introduce vulnerabilidades operativas y de concentración que deben ser tenidas en cuenta por supervisores y entidades.

En suma, la IA ofrece oportunidades claras para mejorar la eficiencia y la capacidad predictiva en los mercados, pero su adopción debe ir acompañada de una gobernanza robusta, salvaguardas regulatorias y un uso responsable que preserve

la estabilidad financiera. El reto no consiste solo en explotar la potencia de la herramienta, sino en hacerlo de manera que el progreso técnico se traduzca en beneficios sostenibles para los inversores y para el sistema en su conjunto.

BIBLIOGRAFÍA

- Cochrane, J. H. (2005). *Asset Pricing* (2.ª ed.). Princeton, Estados Unidos: Princeton University Press.
- Cornell, B. (2020). Medallion Fund: The Ultimate Counterexample? (December 16, 2019). *Journal of Portfolio Management*, 46(4), 2020, <https://jpm.pm-research.com/content/46/4/156>, Available at SSRN: <https://ssrn.com/abstract=3504766> or <http://dx.doi.org/10.2139/ssrn.3504766>.
- Fama, E. F., y French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1), 3-56.
- Guo, S., y Liu, Q. (2024). *Is the annualized compounded return of Medallion over 35 por 100?* Preprint.
- Kelly, B., y Xiu, D. (2023). Financial machine learning. *Foundations and Trends® in Finance*, 13(3-4), 205-363.
- Ross, S. A. (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, 13(3), 341-360.
- Bryzgalova, S., Pelger, M., y Zhu, J. (2023). Forest through the trees: Building cross-sections of stock returns. *Review of Financial Studies*, 36(1), 1-42.
- Gu, S., Kelly, B., y Xiu, D. (2020). Empirical asset pricing via machine learning. *Review of Financial Studies*, 33(5), 2223-2273.
- Harvey, C., Liu, Y., y Zhu, H. (2016). ...and the cross-section of expected returns. *Review of Financial Studies*, 29(1), 5-68.
- Fama, E. F., y French, K. R. (1998). Value versus growth: The international evidence. *Journal of Finance*, 53(6), 1975-1999.
- Hou, K., Van Dijk, M. A., y Zhang, Y. (2012). The implied cost of capital: A new approach. *Journal of Accounting and Economics*, 53(3), 504-526.

Lakonishok, J., Shleifer, A., y Vishny, R. (1994). Contrarian investment, extrapolation, and risk. *Journal of Finance*, 49(5), 1541-1578.

Abarbanell, J. S., y Bernard, V. L. (1992). Tests of analysts' overreaction/underreaction to earnings information as an explanation for anomalous stock price behavior. *Journal of Finance*, 47(3), 1181-1207.

Chan, L. K. C., Jegadeesh, N., y Lakonishok, J. (1996). Momentum strategies. *Journal of Finance*, 51(5), 1681-1713.

Givoly, D., y Lakonishok, J. (1979). The information content of financial analysts' forecasts of earnings: Some evidence on semi-strong inefficiency. *Journal of Accounting and Economics*, 1(2), 165-185.

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

López de Prado, Marcos (2018). *Advances in Financial Machine Learning*. Hoboken, New Jersey: Wiley.