

Optimización de cartera de activos financieros aplicando aprendizaje automático

Autor: Antonio Caparrini López

Director: Jordi Escayola Mansilla

Índice

1. Introducción

1. Mercados financieros
2. Rendimiento de una acción
3. SP500

2. Motivación y objetivos

3. Desarrollo

1. Conjunto de datos
2. Factores
3. Algoritmos
4. Backtest

4. Resultados

1. Ranking importancias XGBoost
2. Importancias XGBoost
3. Tasas de acierto
4. Rentabilidades
5. Resumen

5. Conclusiones y trabajo futuro

Introducción

Introducción - Mercados financieros

S&P 500

3798,09
+49,95 (+1,33 %)



Nasdaq

13.032,87
+292,07 (+2,29 %)



NIKKEI 225

27.490,13
+434,19 (+1,60%)



Dólar/Euro

1,2273
-0,0051 (-0,4171 %)



Petróleo Brent

54,46
+0,16 (+0,29%)



- Mercados financieros tienen como objetivo poner en contacto a **oferentes** de capital con **demandantes** de **capital** y además determinar los precios justos de los **activos financieros**.
- **Interesados** en los mercados podemos considerar:
 - **Fondos de inversión**
 - **Fondos de pensiones**
 - **Inversores particulares**

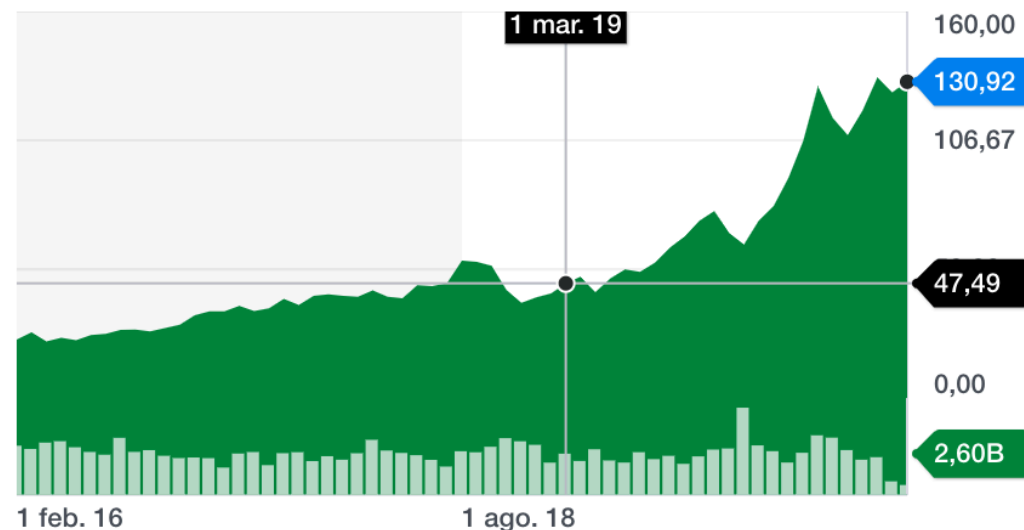


Introducción - Rendimiento de una acción

Apple Inc. (AAPL)

| | | | |
|-------------------------|-----------------|---------------------------------------|-----------------------------|
| Cierre anterior | 126,60 | Capitalización de mercado | 2,193T |
| Abrir | 128,36 | Beta (5 años al mes) | 1,28 |
| Oferta | 130,23 x 1200 | Ratio precio/beneficio (TMTM) | 39,75 |
| Precio de compra | 130,24 x 1200 | BPA (TTM): | 3,28 |
| Rango diario | 127,86 - 131,03 | Fecha de beneficios | 25 ene. 2021 - 31 ene. 2021 |
| Intervalo de 52 semanas | 53,15 - 138,79 | Previsión de rentabilidad y dividendo | 0,82 (0,65%) |
| Volumen | 67.486.542 | Fecha de exdividendo | 05 nov. 2020 |
| Media Volumen | 114.411.661 | Objetivo est 1a | 129,84 |

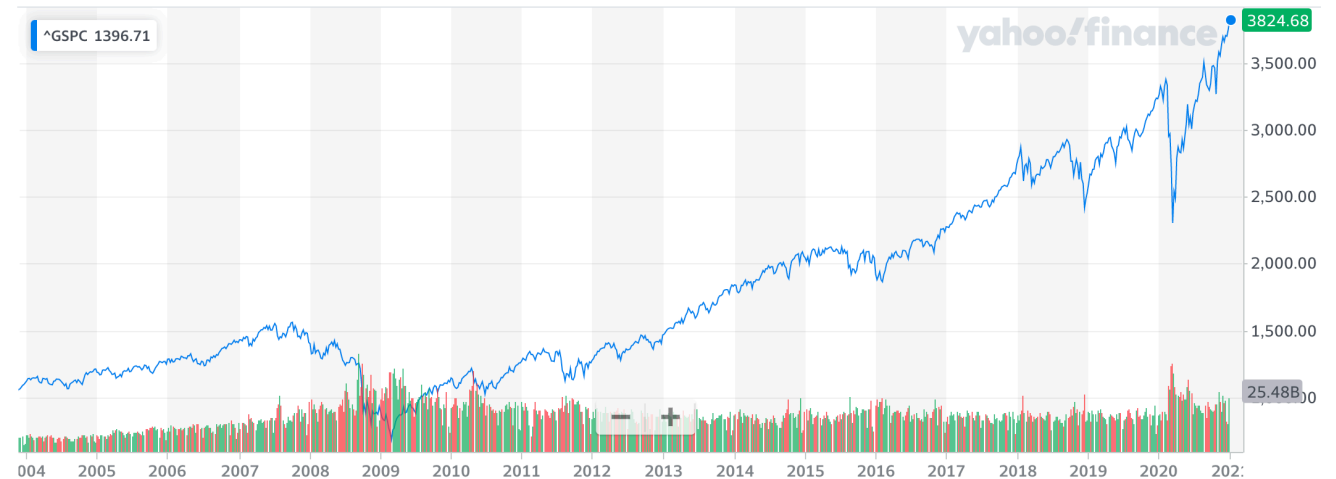
1D 5D 1M 6M YTD 1Y 5Y Max [Pantalla completa](#)



Los precios de las transacciones no se obtienen de todos los mercados

Introducción - SP500

- El índice Standard & Poor's 500 (Standard & Poor's 500 Index), también conocido como S&P 500, es uno de los **índices bursátiles más importantes de Estados Unidos**.
- Se lo considera el **índice más representativo** de la situación real del **mercado**.
- Tiene en cuenta la **capitalización bursátil de 500 grandes empresas** que cotizan en las bolsas NYSE o NASDAQ.
- Captura aproximadamente el **80%** de toda la **capitalización de mercado** en Estados Unidos.



Porcentaje de fondos de inversión de EEUU en renta variable internacional que son superados por sus índices de referencia en 2017

| Categoría Fondos Inversión | Índice de referencia | 1 años | 3 años | 5 años | 10 años |
|------------------------------|----------------------|---------|---------|---------|---------|
| Renta Variable Global | S&P Global 1200 | 50.21 % | 77.45 % | 77.71% | 81.98 % |
| Renta Variable Internacional | S&P 700 | 53.95 % | 69.40 % | 70.93 % | 81.68% |
| Renta Variable Emergentes | S&P/IFCI | 64.89 % | 78.92 % | 77.78 % | 85.14 % |

Motivación y objetivos

Motivación

- Tengo **conocimientos de productos financieros y riesgos** gracias a mi formación y experiencia laboral en entidades financieras.
- Similar a lo empleado en este trabajo, he realizado dos publicaciones utilizando **algoritmos arbóreos** de aprendizaje automático optimizando los hiperparámetros usando **algoritmos genéticos** aplicado a reconocimiento de género musical y riesgo de crédito:
 - [Automatic subgenre classification in an electronic dance music taxonomy](#)
 - [Explainability of a Machine Learning Granting Scoring Model in Peer-to-Peer Lending](#)
- Además, tengo un **interés personal en la bolsa y los mercados**, formándome e invirtiendo como afición personal.

Objetivos

- Utilizar **aprendizaje automático** para construir un modelo que **pronostique activos** con **mayores rendimientos** futuros.
- **Optimizar** los **hiperparámetros** de los algoritmos de aprendizaje automático usando **algoritmos genéticos**.
- Crear un **método de selección de activos** de entre los listados en el índice SP500.
- **Definir la estrategia** para verificar el comportamiento de la estrategia y comprobar si históricamente ha funcionado.

Desarrollo

Desarrollo - Conjunto de datos

Quandl Search for data EXPLORE MONETIZE YOUR DATA

PUBLISHED BY SHARADAR
Core US Fundamentals Data

DATA DOCUMENTATION USAGE

LIST OF TABLES

Core US Fundamentals EXPAND

| ticker | dimension | calendardate | datekey | reportperiod | lastupdated | accoci |
|--------|-----------|--------------|------------|--------------|-------------|-----------|
| ZZ | MRY | 2011-12-31 | 2011-11-27 | 2011-11-27 | 2019-05-24 | -509000 |
| ZZ | MRY | 2010-12-31 | 2010-11-28 | 2010-11-28 | 2019-05-24 | 7005000 |
| ZZ | MRY | 2009-12-31 | 2009-11-29 | 2009-11-29 | 2019-05-24 | -1053000 |
| ZZ | MRY | 2008-12-31 | 2008-11-30 | 2008-11-30 | 2019-05-24 | -19711000 |
| ZZ | MRY | 2007-12-31 | 2007-12-02 | 2007-12-02 | 2019-05-24 | 9239000 |
| ZZ | MRY | 2006-12-31 | 2006-11-26 | 2006-11-26 | 2019-05-24 | 7793000 |
| ZZ | MRY | 2005-12-31 | 2005-11-27 | 2005-11-27 | 2019-05-24 | 2638000 |

DISPLAYING 7 ROWS. SHARADAR/SF1

Tickers and Metadata EXPAND

| table | permaticker | ticker | name | exchange | isdelisted | category |
|-------|-------------|--------|---------------------------|----------|------------|-------------|
| SF1 | 196290 | A | Agilent Technologies I... | NYSE | N | Domestic Co |
| SF1 | 124392 | AA | Alcoa Corp | NYSE | N | Domestic Co |
| SF1 | 122827 | AAAB | Admiralty Bancorp Inc | NASDAQ | Y | Domestic Co |
| SF1 | 120538 | AAAGY | Altana Aktiengesellsc... | NYSE | Y | ADR Commo |
| SF1 | 155760 | AAAP | Advanced Accelerator | NASDAQ | Y | ADR Commo |
| SF1 | 165444 | AABC | Access Anytime Banc... | NASDAQ | Y | Domestic Co |
| SF1 | 119789 | AAC1 | Arcadia Financial Ltd | NYSE | Y | Domestic Co |

DISPLAYING 7 ROWS. SHARADAR/TICKERS

PREMIUM DATA

DESCRIPTION
Up to 20 years of history, and growing, for more than 14,000 companies and 150 indicators. Active & Delisted. Point-in-time. Incl/Excl Restatements.

DELIVERY FREQUENCY Daily

DATA FREQUENCY Annual, Quarterly

REPORTING LAG < 1 day

HISTORY Dec 1997

COVERAGE 14,000+ US companies, 150 indicators

AVAILABILITY Premium

Tables API

Unbookmark this feed

FREE SAMPLE ENABLED

A subset of this data is available for testing without a subscription. To view the full data, you must subscribe to the product.

VIEW PRICING

- **Quandl** ofrece conjuntos de datos de diferentes proveedores
- **Sharadar** es una **firma independiente** de análisis e investigación **especializada** en la extracción, estandarización y organización de **datos financieros** de los informes periódicos presentados por las empresas.
- **Sharadar Core Us Equities Bundle**
- 14.000 empresas de USA desde 1997
- Indicadores, precios y fundamentales.

Desarrollo - Factores

- Los **factores** (o anomalías) son las características que se tienen en cuenta en estudios sobre el retorno de los activos.
- Hemos obtenido o generado **22 factores** para más de 1400 acciones que han pertenecido al **SP500 desde 1997**.
- Para el problema de clasificación generamos la **variable objetivo** en función de los **retornos a futuro de 12 meses** teniendo en cuenta la tabla.
- La selección la realizaremos de los **15 activos** con mayor probabilidad de pertenecer a la **clase 4**.

| Retorno 12 meses | Clase |
|----------------------------|-------|
| $retorno \geq 0,15$ | 4 |
| $0,05 \leq retorno < 0,15$ | 3 |
| $0 \leq retorno < 0,05$ | 2 |
| $-0,15 \leq retorno < 0$ | 1 |
| $retorno < -0,15$ | 0 |

| Factor | Category |
|------------------------------|-------------------|
| 6-month prior return | Momentum |
| 11-month prior return | Momentum |
| Return on equity | Profitability |
| beta 3Y covariance | Intangible |
| beta 3Y linear reg | Intangible |
| Price-to-book | Value-vs-growth |
| Earnings-to-price | Value-vs-growth |
| Price-to-sales | Value-vs-growth |
| Market capitalization | Trading frictions |
| Enterprise-value | Value-vs-growth |
| evebit | Value |
| evebitda | Value |
| 12 month lagged return | Intangibles |
| 24 month lagged return | Intangibles |
| Operating cash flow-to-price | Value-vs-growth |
| Investment-to-price | Investment |
| Earning per share | Profitability |
| Current ratio | Intangibles |
| Operating cashflow-to-equity | Profitability |
| Capex | Intangibles |

Desarrollo - Algoritmos



Librería software libre para Python orientada al análisis de datos:

- **Árbol de decisión**
- **Bosque aleatorio**



Publicado en 2014. Popular implementación eficiente de algoritmos de aprendizaje automático usando el paradigma “gradient boosting”.



DISTRIBUTED
EVOLUTIONARY
ALGORITHMS IN
PYTHON

Facilita el “prototipado” rápido de **optimizaciones** con **algoritmos genéticos**.
Métrica objetivo **tasa de aciertos balanceada** en una **validación cruzada de 5 iteraciones**.

Procedimiento a seguir

1. Generamos **parámetros optimizados** con **algoritmo genético** para el **primer clasificador**
2. **Entrenamos clasificadores** en **diferentes fechas** para **seleccionar activos**
3. **Obtenemos el retorno** de los activos por el **periodo** que los hemos **mantenido** en cartera

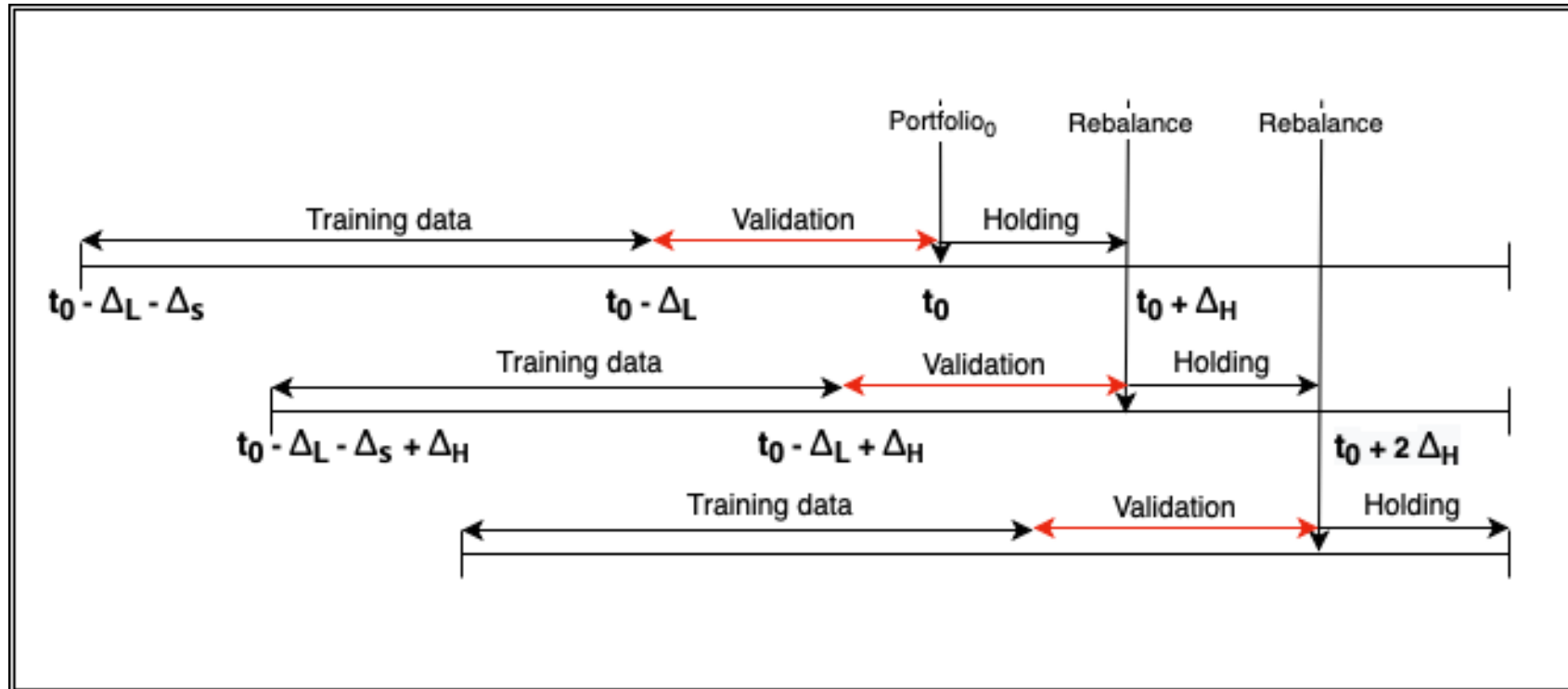
Desarrollo - Backtest

Backtest se define como una prueba de un modelo utilizando datos históricos. **Replicamos el comportamiento** que utilizaríamos de aplicar la estrategia a futuro pero **desde una fecha pasada** para verificar los resultados.

Para ello vamos a contar con las siguientes definiciones:

- ΔS : **Rango** de tiempo del **conjunto de datos** utilizado para **entrenar** el modelo (3 años)
- ΔH : Periodo de **mantenimiento de los activos en cartera** entre fechas de rebalanceo (4 meses)
- ΔL : Periodo de **datos no utilizados** para evitar sesgos con la variable objetivo (1 año)
- t_0 : **Primera** generación de la **cartera** con suficientes datos (3 de enero de 2005).

Desarrollo - Backtest



- $t_0 - \Delta_L - \Delta_S$: Fecha donde se **inicia** el **conjunto** de datos de **entrenamiento**
- $t_0 - \Delta_L$: Fecha **final** del **conjunto** de datos de **entrenamiento** y fecha donde **empieza** el **conjunto de validación**
- $t_0 + \Delta_H$: **Fecha** a comprobar los **rendimientos** de los activos seleccionados en t_0

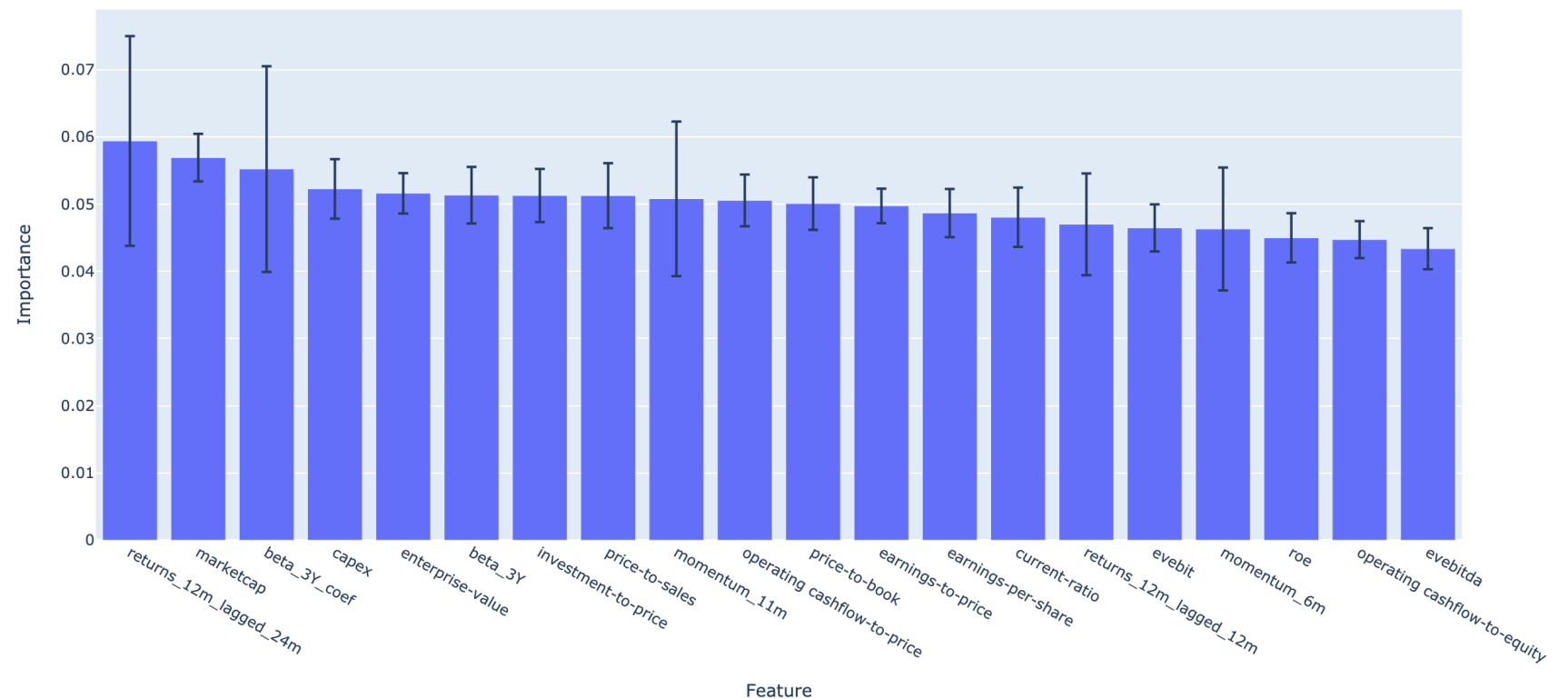
Resultados

Resultados - Ranking importancias XGBoost

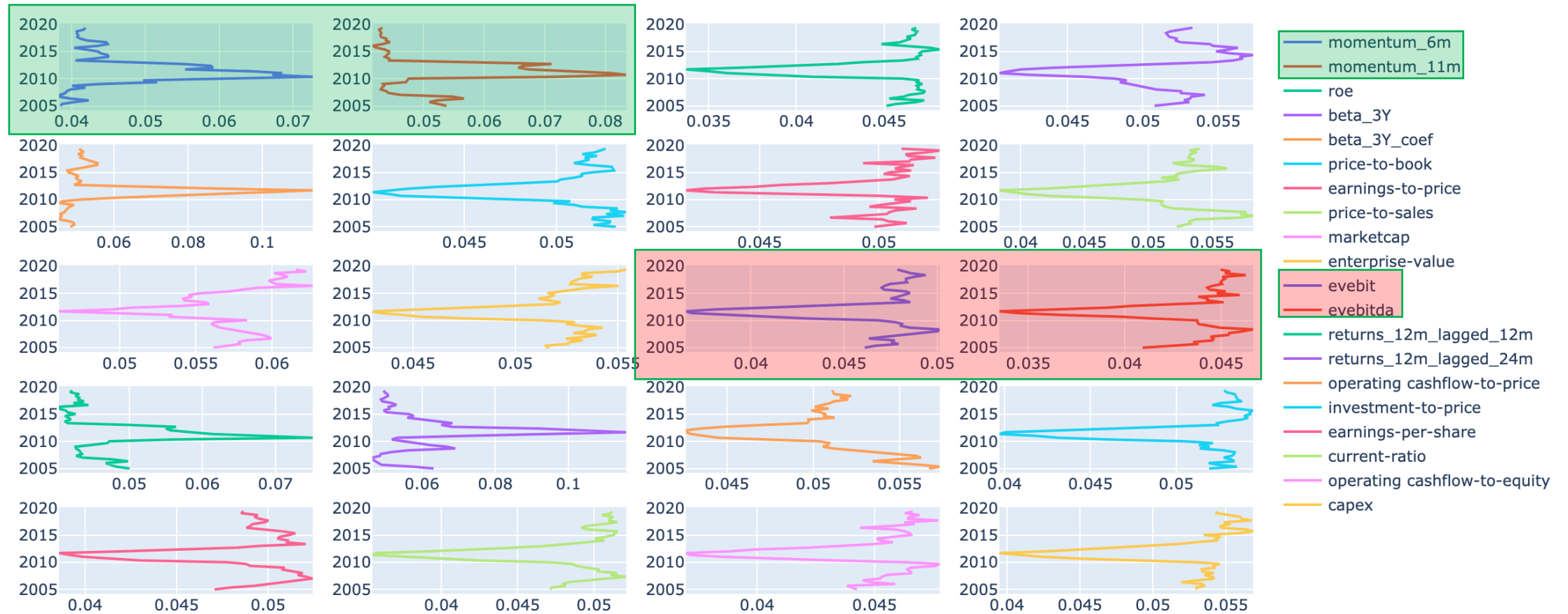
La **importancia** de las **características** se obtiene de XGBoost y es un valor que representa la relación del **número de veces** que la **característica es usada para dividir datos**.

En **gráfico** muestra la **media** y la **desviación** de las **características de todos los clasificadores**.

Las **características** tienen una **importancia similar** pero las más importantes cambian como se puede ver en las **altas desviaciones**.

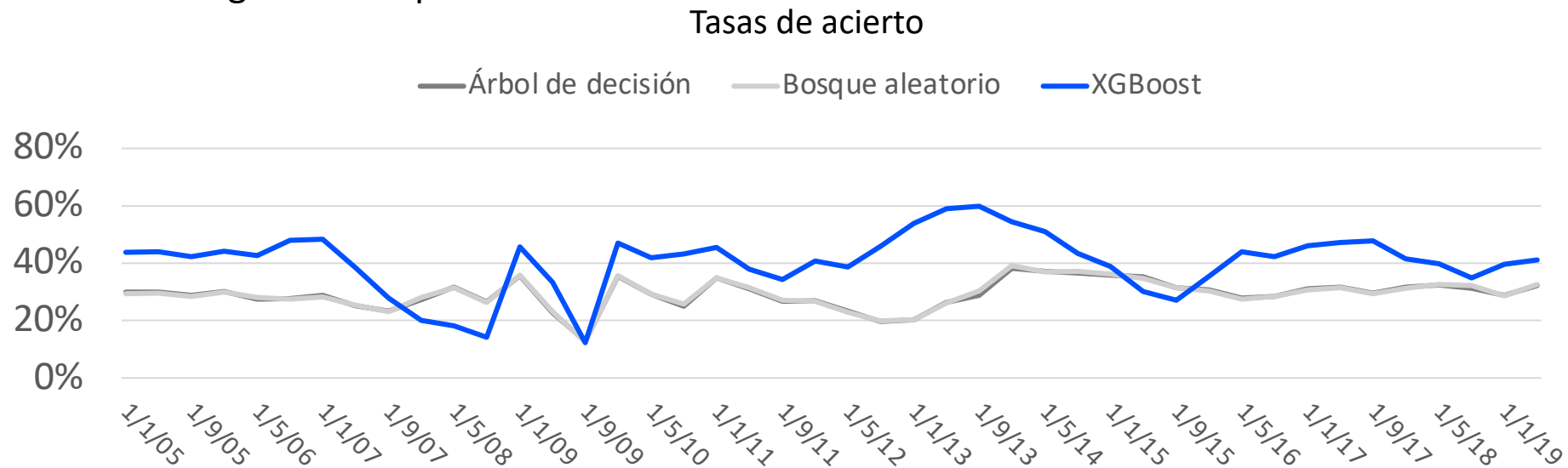


Resultados - Importancias XGBoost



Resultados - Tasas de acierto

En el gráfico se muestran las diferentes **tasas de acierto (predicciones correctas / predicciones totales)** de los clasificadores a lo largo del tiempo:

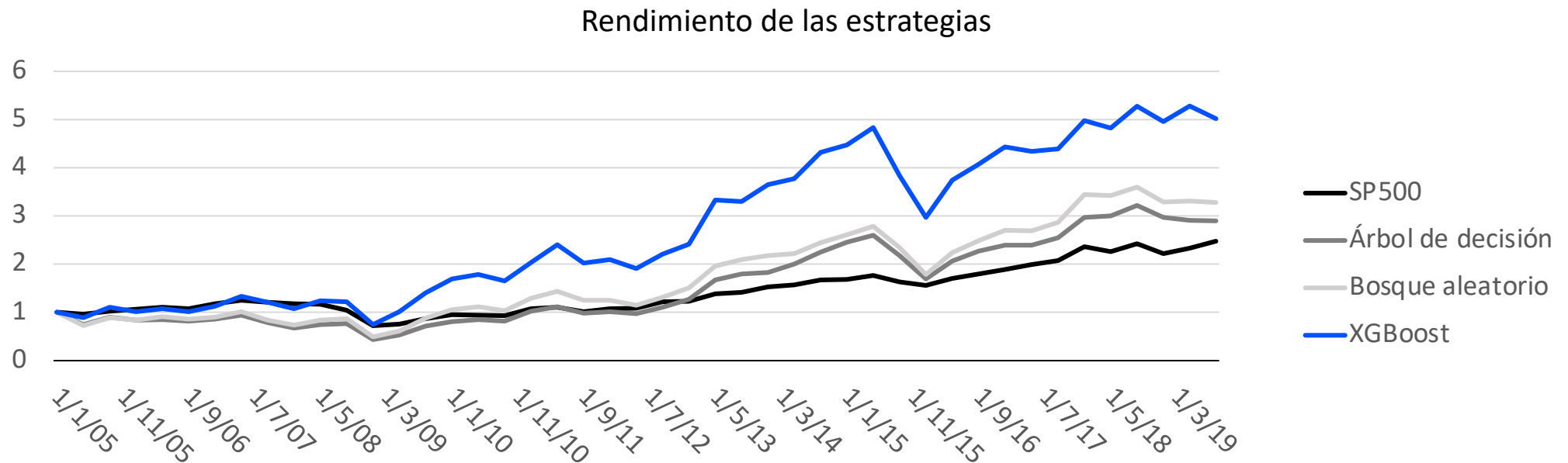


Árbol de decisión y bosque aleatorio tienen tasas similares y hay una **tasa de acierto** sistemáticamente **mejor** en el **XGBoost**.

Sin embargo, esta **tasa de acierto** es sobre **todo el conjunto de validación** y **no refleja los rendimientos sobre los 15 activos** con más probabilidad de rentabilidad que seleccionamos.

Resultados - Rentabilidades

Cada **cartera teórica** “modelo” comienza el 3 de enero de 2005 con un valor de 1 (podemos considerar **1\$**).



La estrategia utilizando **XGBoost bate el índice de referencia** y tiene una tendencia histórica mayor.

Resultados - Resumen

| Estrategia | Retornos totales | Retornos anualizados | Volatilidad anual |
|-------------------|------------------|----------------------|-------------------|
| SP500 | 147,63% | 6,34% | 17,81% |
| Árbol de decisión | 189,63% | 7,48% | 33,54% |
| Bosque aleatorio | 228,20% | 8,39% | 38,70% |
| XGBoost | 402,16% | 11,56% | 39,42% |

- La rentabilidad de las estrategias con **árbol de decisión y bosque aleatorio** producen **ligeramente mayor rentabilidad** pero a cambio de una **volatilidad mucho mayor** que el mercado.
- El resultado con la estrategia de **XGBoost** produce una **rentabilidad mayor** (más del doble sobre el periodo de backtest) pero a cambio de **más del doble** de volatilidad que el mercado.
- El **backtest tiene limitaciones**:
 - **No se han considerado gastos y comisiones** por operaciones e intermediarios financieros.
 - Actuar sobre el mercado, **afecta al mercado**.
 - **Rentabilidades pasadas no garantizan rentabilidades futuras**.

Conclusiones y trabajo futuro

Conclusiones

- Se ha **construido** un **modelo de selección de activos** basado en algoritmos de **aprendizaje automático optimizados** usando **algoritmos genéticos**.
- Entrenar el algoritmo con datos recientes obtiene buenos resultados y se percibe como la **importancia de las características** demuestra que en **distintos momentos** los factores más relevantes **cambian**.
- Mediante el **backtest** hemos verificado que la **selección de activos** mediante el planteamiento propuesto ofrece una **rentabilidad superior** que la del **mercado**.
- Es necesario tener en cuenta que la estrategia se desarrolla a **largo plazo** y que en **algunos instantes** del tiempo la **estrategia** tiene **rendimientos negativos** y una **mayor volatilidad**.

Trabajo futuro

- **Ampliar la cantidad de factores** utilizados o hacer una preselección previa para cada modelo.
- Aplicar el procedimiento a **otro universo de activos** (por ejemplo FTSE 100, IBEX 35 o NASDAQ 100).
- Debido al cambio de factores importantes en función del contexto económico **combinar los modelos generados con modelos que pronostiquen el contexto económico.**
- Aplicar otras técnicas como **redes neuronales.**
- Investigar **explicabilidad de los algoritmos** que es un campo de interés para inversores y reguladores.

Disclaimer



El contenido de este trabajo **no constituye una oferta o recomendación de compra o venta de instrumentos financieros.**

El inversor debe ser consciente de que los valores e instrumentos financieros a que se refieren pueden no ser adecuados a sus objetivos concretos de inversión, por lo que el inversor debe adoptar sus propias decisiones de inversión, procurándose a tal fin el asesoramiento especializado que considere necesario.

El inversor debe tener en cuenta que la evolución pasada de los valores e instrumentos o los **resultados pasados de las inversiones no garantizan la evolución o resultados futuros.**



Muchas gracias