



Universitat
de les Illes Balears

TRABAJO DE FIN DE MÁSTER

AUTOMATIZACIÓN DE UNA OPERATIVA DE TRADING USANDO APRENDIZAJE AUTOMÁTICO

Marcos Muelas Tenorio

Máster Universitario en Sistemas Inteligentes (MUSI)

Especialidad: Inteligencia Artificial

Centro de Estudios de Postgrado

Año académico 2022-23

AUTOMATIZACIÓN DE UNA OPERATIVA DE TRADING USANDO APRENDIZAJE AUTOMÁTICO

Marcos Muelas Tenorio

**Trabajo de Fin de Máster
Centro de Estudios de Posgrado
Universidad de las Illes Balears**

Año académico 2022-23

Palabras clave:

Trading Cuantitativo, Aprendizaje Automático, Inversión Financiera.

Nombre Tutor/Tutora: Isaac Lera Castro

Nombre Tutor/Tutora: Javier Varona Gomez

Automatización de una operativa de trading usando aprendizaje automático

Marcos Muelas Tenorio

Tutor: Isaac Lera Castro y Javier Varona Gomez

Trabajo de fin de Máster Universitario en Sistemas Inteligentes (MUSTI)

Universitat de les Illes Balears

07122 Palma, Illes Balears, Espanya

marcos.muelas1@estudiant.uib.cat

Resumen—El presente trabajo introduce el concepto de trading algorítmico, destacando su origen en la observación de precios y la representación gráfica de la evolución de activos. Se hace incapié en el beneficio existente de un enfoque sistemático y la transición hacia estrategias cuantitativas descritas como un sistema automatizado basado en modelos matemáticos y se destaca algunos de los avances en computación inteligente y algoritmos de aprendizaje automático para la predicción bursátil, con un enfoque en la adaptación de modelos a la variabilidad del mercado.

En este contexto, se plantea una estrategia de trading cuantitativo con el objetivo de predecir el precio de cierre diario en función de diversas variables y se plantea el modelado sobre la acción de Microsoft. Se simulan operaciones, suponiendo un caso ideal y uno real, con diferentes modelos con objetivo de demostrar su capacidad para hacer predicciones precisas y rentables. En este estudio se reflexiona sobre la importancia de incorporar métodos de aprendizaje automático en la automatización de toma de decisiones y la optimización de estrategias de inversión.

ABSTRACT

This paper introduces the concept of algorithmic trading, highlighting its origin in the observation of prices and the graphical representation of the evolution of assets. It emphasizes the existing benefit of a systematic approach and the transition towards quantitative strategies described as an automated system based on mathematical models and highlights some of the advances in intelligent computing and machine learning algorithms for stock market prediction, with a focus on adapting models to market variability. In this context, a quantitative trading strategy is proposed with the objective of predicting the daily closing price as a function of several variables and modeling is proposed on Microsoft stock. Trades are simulated, assuming an ideal and a real case, with different models with the aim of demonstrating their ability to make accurate and useful predictions. The study aims to highlight the importance of the implementation of artificial intelligence in the automation of decision making and the optimization of investment strategies.

Palabras clave: Trading Cuantitativo, Aprendizaje Automático, Inversión Financiera.

I. MOTIVACIÓN

La automatización ofrece la capacidad de ejecutar estrategias de manera sistemática y libre de sesgos

emocionales, aprovechando algoritmos y modelos matemáticos para analizar datos y tomar decisiones en tiempo real. Esto no solo puede mejorar la precisión de las operaciones, sino también reducir la carga de trabajo manual. Además, la automatización permite aprovechar oportunidades de mercado que podrían pasar desapercibidas para un operador humano y proporciona la posibilidad de gestionar carteras de manera más efectiva. En resumen, la búsqueda de estrategias rentables y su automatización busca mejorar la eficiencia, minimizar errores humanos y maximizar el rendimiento en el dinámico y competitivo entorno financiero.

II. INTRODUCCIÓN

Antiguamente, la mayoría de los inversores decidían la operativa que ejecutarían en el mercado por intuición, otros, utilizaban las noticias publicadas en los periódicos, o buscaban obtener información privilegiada. Posteriormente, surgieron las primeras personas que comenzaron a observar los precios de los activos y a representar su evolución mediante gráficos, dando como origen a todo el movimiento del análisis técnico, lo que propicio a su vez la aparición de numerosos indicadores basados en cálculos aplicados sobre dichos datos.

La información existente era cada vez más abundante, pero la capacidad de procesarla aun era mínima lo que resultaba en análisis discrecionales, llegando a conclusiones distintas según la persona que lo realizaba. La necesidad de establecer un procedimiento sistemático para evaluar los precios utilizando parámetros y objetivos concretos empezó a llamar la atención de personas relacionadas con las ciencias, matemáticas e ingeniería, familiarizada con la ejecución metódica a la hora de poner a prueba una hipótesis, analizando e interpretando los resultados de los experimentos.

Esta fue la semilla de lo que acabaría convirtiéndose en lo que hoy conocemos como trading algorítmico. Es simplemente una aproximación a los mercados basada en el análisis de la volatilidad del precio aplicando métodos sistemáticos, secuenciales, matemáticos y estadísticos fruto de una estrategia determinada [2]. El objetivo es trabajar con probabilidades recogidas en los modelos, los cuales serán más efectivos que cualquier comportamiento humano debido a que

estarán libres de sesgos cognitivos y juicios discrecionales. Como menciona Ray Dalio, fundador y codirector de inversiones de Bridgewater Associates y destacado inversor macro a nivel global, en su libro titulado *«Principios»* [3] :

«Creo que nada es totalmente cierto. Creo que lo mejor que podemos esperar es que sea altamente probable»

Dado que el alcance total es muy amplio, en el presente trabajo se opta por abordar el estudio del comportamiento de algunos de los diferentes modelos existentes con el fin de extraer conclusiones propias y corroborar las obtenidas tras la investigación realizada.

Los modelos seleccionados se clasifican dentro del machine learning tradicional y son los siguientes:

- Regresión Lineal.
- Regresión Polinómica.
- Máquinas de soporte vectorial.
- Árboles de decisión.
- Bosques aleatorios.

Para este estudio, los datos utilizados en los modelos se componen de los datos históricos de la volatilidad del precio de la acción de Microsoft, junto con una batería de indicadores técnicos que serán desarrollados en secciones posteriores.

En referencia a la operativa de trading se plantea una estrategia de tipo temporal, en la que se abre una posición de compra/venta en función de si la predicción del precio obtenida es mayor/menor que el precio de apertura de la sesión de negociación, con objetivo de observar los resultados obtenidos para cada uno de los casos supuestos, uno ideal (en ausencia de spread y comisiones) y otro real (teniendo en cuenta el broker seleccionado y los márgenes de compra/venta).

III. ESTADO DEL ARTE

La inversión financiera se entiende como un conjunto de metas, intenciones o propósitos que un inversor tiene en mente a la hora de gestionar una cartera de activos financieros con el objetivo de obtener un beneficio sostenible en el tiempo. Cuando se realiza una inversión, implícitamente se asume un costo de oportunidad al renunciar a esos recursos en el presente o por un periodo determinado de tiempo para lograr obtener un beneficio a futuro.

En líneas generales, los objetivos de inversión se fundamentan en la seguridad, el crecimiento y los ingresos. Una buena definición de los objetivos ayuda a tomar mejores decisiones en función de los intereses buscados y la metodología que se quiera aplicar.

En la práctica, la predicción bursátil puede realizarse mediante análisis fundamental y análisis técnico. El análisis fundamental es el uso más convencional, que trata de determinar el valor o el precio de una acción basándose en estados financieros como la cuenta de resultados, el balance y el estado de flujos de caja. En otras palabras, el objetivo principal del análisis fundamental es estimar el valor intrínseco de una empresa ya que tiene una correlación positiva y significativa con los resultados futuros. Es la investigación previa para la inversión en valor conocida también como inversión a largo plazo. En

cambio, el análisis técnico se realiza basándose en gráficos e indicadores técnicos basados en datos históricos y suele utilizarse para predecir horizontes temporales de corto a medio plazo [1].

Se define portfolio como una cartera de activos de inversión y su gestión está basado en un proceso de toma de decisiones de inversión fundamentado en estrategias personalizadas con objetivo de maximizar el rendimiento para un horizonte temporal.

Existen dos enfoques populares para gestionar la cartera de inversión, el tradicional y el cuantitativo. Ambos coinciden en varios términos, como la investigación de un pequeño conjunto de factores impulsores clave de los activos objetivo, el análisis de datos histórico para estimar estos impulsores clave, la adopción de criterios de elegibilidad para las decisiones de selección de valores y la evaluación del rendimiento a lo largo del tiempo. Sin embargo, mientras que la gestión tradicional de carteras se centra en el análisis en profundidad de juicios, cambios de régimen, características clave y factores cualitativos, la gestión de cartera cuantitativa se centra en la exploración del universo, la disciplina, verificación y gestión del riesgo [10].

El trading cuantitativo consiste en estrategias de negociación basadas en modelos matemáticos para diseñar un sistema de trading automatizado. El desarrollo de un modelo de predicción preciso es el proceso más crítico para construir una cartera eficiente en el enfoque cuantitativo y no es raro que se considere una de las tareas con mayores dificultades del mercado financiero debido a la complejidad de los atributos de las series temporales multivariantes y a la cantidad de datos financieros implicados [6].

El primer paso para elaborar la estrategia puede comenzar de dos formas distintas:

- Formulación de una hipótesis que describa el comportamiento que se desea estudiar.
- Investigar comportamientos no evidentes utilizando técnicas de minería de datos.

El segundo paso consiste en crear un experimento donde sea posible simular la estrategia planteada y observar los resultados para poder analizarlo en profundidad. Posteriormente se recogen los datos que puedan verificar o desmentir la validez de la estrategia planteada, será imprescindible que los datos se recojan de manera objetiva y sin ninguna inferencia por parte del observador. En última instancia se procede al análisis de los datos recogidos, de tal forma que se ha de llegar a la conclusión de decir si la estrategia planteada es correcta, o por el contrario si hay que replantearla o descartarla [9].

El uso de un método científico a la hora de crear modelos para operar en los mercados hará llegar a conclusiones objetivas, veraces y robustas, de forma que las decisiones tomadas serán resultado de un proceso riguroso y metódico fundamentado por datos empíricos.

El objetivo es analizar el comportamiento del precio y buscar comportamientos globales que se repitan de manera

constante. De esta forma, se definen dos tipos principales de estrategias: las estrategias basadas en el propio precio y aquellas con inferencias técnicas en el mercado.

Es importante resaltar que, si bien la parte de la entrada al mercado es importante, la parte de salida es absolutamente fundamental. Una vez estes dentro del mercado no se sabrá a ciencia cierta cómo evolucionará el precio, solo mediante la gestión de la posición se podrá mantener las probabilidades de éxito frente a la constante incertidumbre [9].

III-A. Trading Cuantitativo

En los últimos años, los investigadores han construido principalmente modelos estadísticos para describir las series temporales de los precios de las acciones y las cotizaciones bursátiles. Cabe señalar que los métodos de computación inteligente representados por algoritmos de aprendizaje automático también presentan un vigoroso avance en la predicción bursátil con el desarrollo de la inteligencia artificial.

El trading cuantitativo es un sistema automatizado o semiautomatizado en el que las estrategias y decisiones de trading se llevan a cabo mediante un conjunto de modelos matemáticos. La idea de la negociación cuantitativa está diseñada para aprovechar las matemáticas estadísticas, los algoritmos informáticos y los recursos computacionales de los sistemas de negociación de alta frecuencia, cuyo objetivo es minimizar el riesgo y maximizar la rentabilidad basándose en el rendimiento de las estrategias.

La primera etapa fundamental del sistema es el proceso de recopilación de datos, en el que los datos pueden obtenerse de fuentes externas, de un proveedor de datos o de una investigación propia [5]. El preprocesamiento y la limpieza de datos son las principales tareas para obtener fuentes de datos fiables. La función del proceso de modelado se centra principalmente en la creación de modelos precisos de predicción, análisis estadístico y optimización. Finalmente, los resultados del análisis se procesan y se convierten en los criterios para la toma de decisiones de inversión. Las dos últimas etapas: modelización y análisis, son empleadas en un proceso iterativo de evaluación de tendencias, determinación de estrategias, backtesting y evaluación del rendimiento de la cartera.

En [8] los autores exponen una comparativa de diferentes sistemas de predicción sobre 609 diferentes activos financieros del mercado chino y americano en dos escenarios diferentes: considerando y no considerando los costes de transacción. Para la implementación de los sistemas de predicción utilizan 6 algoritmos tradicionales de machine learning que son Support Vector Machine, Random Forest, Logistic Regression, Naive Bayes, Árboles de clasificación y regresión y XGBoost (Extreme Gradient Boosting) y 6 modelos utilizando redes DNN que son ampliamente utilizados en el texto y la traducción de voz, como el Perceptrón Multicapa (MLP), Redes Neuronales Profundas (DBN), Autocodificadores Apilados (SAE), Redes Neuronales

Recurrentes (RNN), Memoria A Largo-Corto Plazo (LSTM) y la Unidad Recurrente Controlada (GRU) para entrenar y predecir las tendencias del precio de las acciones basándose en los indicadores técnicos. Las conclusiones presentadas demuestran que las redes DNN se adaptan mejor a las variaciones en los costes de transacción pese a que su rendimiento de un resultado inferior al de los algoritmos de machine learning tradicionales en un escenario ideal en ausencia de comisiones.

En otros trabajos como en [11], Wang y Leu presentan un sistema de predicción del mercado bursátil utilizando un modelo no lineal de media móvil integrada autorregresiva (ARIMA). Los autores obtuvieron resultados eficaces en la predicción del mercado de valores utilizando una función de autocorrelación para verificar si una serie temporal es estacionaria o no estacionaria con el fin de eliminar el ruido. Los autores implementan una red recurrente basada en ARIMA para la predicción bursátil y los resultados experimentales confirman que el método propuesto parece bastante bueno, aunque la precisión de la predicción puede mejorarse filtrando las series de cotizaciones bursátiles para eliminar los comportamientos no estacionarios.

Diebold [4] proponen un enfoque de Algoritmos Genéticos (AG) para predecir un índice bursátil. Básicamente se centran en la discretización de características para reducir la complejidad del espacio. A pesar de presentar algunas limitaciones, el enfoque parece muy prometedor. Una de ellas es que el rendimiento del modelo varía en función del número de elementos de las capas ocultas. Además, los pesos y los umbrales deben determinarse con precisión para evitar el problema de sobreajuste de la RNA.

Finalmente, en los últimos años, las redes LSTM han ganado mucho crédito debido a su peculiaridad, que permite predecir series temporales partiendo de los valores anteriores de la serie. En otra implementación de un LSTM en [7], los autores proponen un sistema GRU que pretende mejorar la eficiencia intentando solucionar los problemas de gradientes de fuga. Las conclusiones resultan muy prometedoras ya que muestra rentabilidades reales sobre el mercado, en este caso supera el rendimiento del índice bursátil SP500 operando sobre distintos portafolios con activos que cotizan en dicho índice durante periodos de tiempo de variada amplitud.

En este contexto se pone en evidencia el creciente interés en el estudio de este campo, ya que la posibilidad de encontrar un modelo que automatice una estrategia de inversión rentable es muy atractiva. Este interés se ve respaldado por la evolución de las tecnologías de computación y la evolución de los algoritmos de aprendizaje automático y redes neuronales en la predicción bursátil. A continuación, se desarrolla el proceso de modelado de la estrategia de inversión, que consiste en la obtención del dataset de entrada, el entrenamiento y evaluación de los modelos propuestos, y la simulación de la operativa, delineando así la integración de la teoría con la aplicación práctica en el contexto de la inversión financiera y el trading cuantitativo.

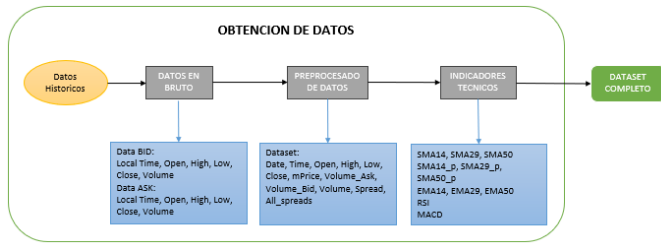


Figura 1: Diagrama de flujo del procesado de datos

IV. DATASET

IV-A. Obtención de datos

La obtención de datos y el preprocesado se divide en distintas etapas y procesos como podemos ver a continuación en la Figura 1 en el diagrama de flujo.

Un factor muy influyente en la precisión de los modelos será la calidad de los datos utilizados para su entrenamiento. Si los datos están contaminados con errores, incompletos o imprecisos, el modelo resultante será propenso a inexactitudes y sus predicciones podrían ser poco fiables. Además, la capacidad de generalización del modelo está estrechamente vinculada a la diversidad y representatividad de los datos. Un modelo bien entrenado debe poder adaptarse a nuevos datos que no haya visto durante el entrenamiento. Por lo tanto, contar con datos variados y representativos permite al modelo aprender patrones más robustos y evitar el sobreajuste, mejorando su capacidad para enfrentar situaciones diversas [4]

En definitiva, la obtención de datos de calidad es esencial para lograr modelos precisos, confiables y efectivos, por lo que para el caso de estudio se decide partir de los datos proporcionados por Dukascopy bank ¹, un banco en línea y broker de divisas que se estableció en 2004 en Suiza, ya que provee de los datos históricos de la variación de precio de multitud de activos. Para el caso propuesto, contamos con la información de la variación de precio de Microsoft desde 2017 hasta la actualidad.

IV-B. Datos en bruto

Los datos brutos se obtienen descargados directamente a través de su web en el apartado de exportación de datos históricos y se descarga en formato .csv. Los datos se componen de los precios de las ordenes de compra y de venta ejecutadas durante el periodo estudiado. Los parámetros extraídos son:

- Local time: Fecha y hora de los datos.
- Open: Precio de apertura del activo durante la temporalidad determinada.
- Close: Precio de cierre del activo durante la temporalidad determinada.
- High: Precio máximo del activo durante la temporalidad determinada.

- Low: Precio mínimo del activo durante la temporalidad determinada.
- Volume: Volumen negociado del activo durante la temporalidad determinada.

IV-C. Preprocesado de datos

Con el fin de englobar toda la información de ambas partes (oferta y demanda), para nuestro caso se crea un único dataset final de 671 muestras que contendrá los siguientes parámetros:

- Date: Fecha del día de cotización.
- Time: Hora del momento de cotización.
- Open: Media (oferta/demanda) de precio de apertura del periodo.
- High: Media (oferta/demanda) de precio máximo del periodo.
- Low: Media (oferta/demanda) de precio mínimo del periodo.
- Close: Media (oferta/demanda) de precio de cierre del periodo.
- mPrice: Media de los 4 precios anteriores (open, high, low, close).
- Volume_Ask: Volumen de demanda.
- Volume_Bid: Volumen de oferta.
- Volume: Suma del volumen total.
- Spread: Promedio de diferencia porcentual entre el precio de oferta y demanda de los 4 precios (open, high, low, close).
- All_spreads: Promedio de la diferencia porcentual entre el precio de oferta y demanda de cada uno de los precios dados (open, high, low, close).

IV-D. Indicadores técnicos

Un indicador técnico es una herramienta utilizada en el análisis técnico de los mercados financieros. Se trata de fórmulas matemáticas y estadísticas que se aplican a los datos históricos de precios y volúmenes de activos financieros. El propósito principal de los indicadores técnicos es proporcionar señales o patrones que ayuden a los traders e inversores a identificar tendencias, momentos de entrada y salida, así como posibles niveles de sobrecompra o sobreventa en un activo. Estos indicadores son utilizados para tomar decisiones de trading y estrategias de inversión basadas en el análisis histórico del comportamiento del precio del activo en el mercado y se desarrollan con mayor detalle en [9]. Para el caso de estudio, se utilizarán los siguientes indicadores técnicos a la hora de elaborar el construir el dataset de entrada:

IV-D1. Media móvil: La media móvil es un cálculo del promedio de ciertos valores de una serie utilizado para analizar un conjunto de datos. Son una serie de valores el cual cada uno corresponde con el promedio de un subconjunto de los datos originales, de esta forma conseguimos una curva suavizada y eliminamos el ruido de las fluctuaciones en periodos de tiempo más cortos.

¹<https://www.dukascopy.com/swiss/spanish/home/>



Figura 2: Evolución del precio MSFT/USD

IV-D2. Media móvil ponderada: Este tipo de media móvil es similar a la anterior, con la diferencia de que asigna mayor peso a los datos más recientes, es decir, da mayor importancia a los valores en el cálculo a los valores más recientes.

IV-D3. Media móvil exponencial: En esta media móvil se aplica un multiplicador determinado por el número de periodos considerados. El cálculo de la media móvil exponencial favorece a los precios más recientes, ya que les da una ponderación mayor.

IV-D4. RSI: El índice de fuerza relativa es un indicador usado en análisis técnico que utiliza la comparación de los movimientos individuales al alza o a la baja de la volatilidad del precio de cierre en un periodo determinado para mostrar la fuerza del precio. Básicamente mide la fuerza con la que suben o bajan los precios de un activo.

IV-D5. MACD: Este indicador mide la distancia de dos medias móviles a lo largo del tiempo, recoge información de diferentes medias móviles relativo a su convergencia y divergencia.

Los periodos seleccionados son los más comunmente utilizados y se trata de 14, 29 y 50 días para las medias móviles, 14 días para el RSI y 14 y 26 días para el MACD [9].

IV-E. Estudio de la evolución del precio

Disponemos de los datos históricos de la volatilidad del precio desde enero de 2017 hasta agosto de 2023, por lo que se realiza un análisis descriptivo anual para dicho periodo y así comparar las distribuciones de la evolución del precio. La variación del precio durante este periodo de puede observar en la Figura 2.

El objetivo de este análisis es proporcionar una comprensión profunda de la volatilidad anual, identificar cambios significativos y detecta outliers en los distintos periodos. La comparación de distribuciones y la validación temporal

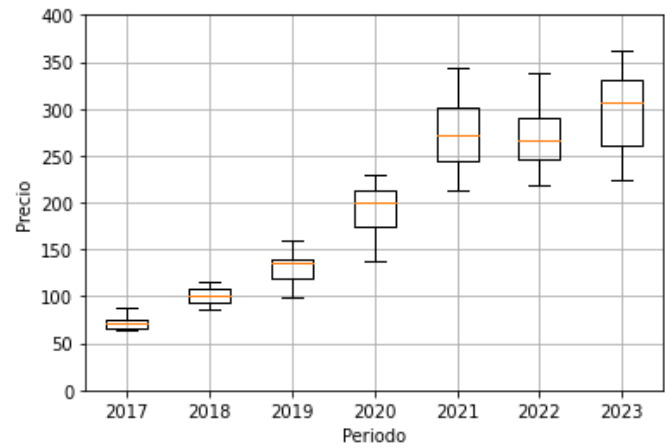


Figura 3: Distribución de precios MSFT/USD

aseguran la elección de años comparables, además, establece rangos de variación para mejorar la generalización del modelo. Las métricas estadísticas calculadas se pueden observar en el Cuadro I.

Métricas	2017	2018	2019	2020	2021	2022	2023
Media	73	101	130	193	275	269	296
Mediana	72	101	135	200	273	266	307
Cuartil25	66	94	119	174	244	246	260
Cuartil50	72	101	135	200	273	266	307
Cuartil75	76	108	140	213	301	290	331
Amplitud	24	29	60	92	132	120	138
Varianza	46	63	227	535	1364	694	1362
Desv. Típica	7	8	15	23	37	26	37

Cuadro I: Analisis descriptivo del precio MSFT/USD

En la Figura 3 se observan las distintas distribuciones de precios en periodos anuales.

De este análisis, usaremos los tres últimos años por tener un patrón parecido y no tan distante a los años pre-covid y covid, por lo que se utilizan los años 2021 y 2022 de entrenamiento y 2023 de testeo.

IV-F. Correlación de variables

La correlación de variables desempeña un papel crucial en la elaboración de modelos de aprendizaje automático para predecir el precio de acciones. Una alta correlación entre características permite al modelo capturar relaciones significativas, mejorando su capacidad predictiva. La selección cuidadosa de variables basada en su correlación puede optimizar el rendimiento del modelo, asegurando que las características más relevantes influyan en la predicción del precio de las acciones. En la Figura 4 aparecen el cálculos de la correlación entre variables utilizando el coeficiente de correlación de pearson.

A la vista de estos resultados, podemos concluir con que todas las variables tienen una alta correlación con la variable objetivo, incluso demasiada en algún caso como el de Open Price o las medias móviles de mismos periodos, a excepción

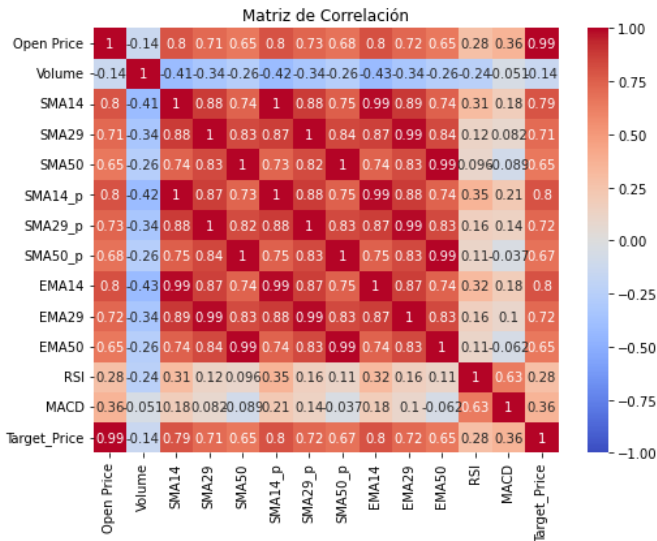


Figura 4: Matriz de correlación



Figura 5: Diagrama de flujo del modelado

del RSI y MACD, cuya correlación es relativa y el volumen que tiene una baja correlación.

V. MODELOS DE PREDICCIÓN

La selección del modelo en aprendizaje automático es la decisión más crucial que puede impactar en el rendimiento, la eficiencia y la interpretación del sistema. Es importante entender las características de los datos, los requisitos del problema y las limitaciones de cada modelo para tomar una decisión informada y obtener los mejores resultados posibles. Los modelos contemplados en el presente trabajo quedan explicados en mayor detalle en [12] y son los que aparecen en el siguiente diagrama de flujo de la Figura 5 y a continuación se exponen:

V-A. Regresión lineal

La regresión lineal es una técnica estadística utilizada para modelar la relación entre una variable dependiente y una o más variables independientes. Se basa en la suposición de que existe una relación lineal entre las variables, es decir, que los cambios en las variables independientes se traducen

en cambios proporcionales en la variable dependiente. El objetivo de la regresión lineal es encontrar la mejor línea recta que se ajuste a los datos observados, minimizando la diferencia entre los valores predichos y los valores reales. Los coeficientes de regresión estimados proporcionan información sobre la magnitud y la dirección de la relación entre las variables.

V-A1. Regresión lineal polinómica: La regresión polinómica es una extensión de la regresión lineal que permite modelar relaciones no lineales entre variables. En lugar de ajustar una línea recta a los datos, la regresión polinómica utiliza una función polinómica para describir la relación entre la variable dependiente y las variables independientes. Esta función polinómica puede ser de grado superior, lo que permite capturar patrones más complejos en los datos y se pueden obtener mejores ajustes a los datos debido a su mayor flexibilidad. Sin embargo, es importante tener en cuenta que el uso de polinomios de grado muy alto puede llevar a un sobreajuste, es decir, un modelo que se ajusta demasiado a los datos de entrenamiento, pero no generaliza bien a nuevos datos.

V-B. Máquinas de soporte vectorial

Los datos de entrenamiento se mapean a un espacio de características de mayor dimensión mediante una función de kernel, similar a la SVM para clasificación. Luego, se busca encontrar una función lineal en este espacio de características que se ajuste a los datos con el menor error posible. Para ello, se definen límites de tolerancia alrededor de los puntos de entrenamiento, permitiendo que algunos puntos se encuentren dentro de estos límites sin afectar significativamente la función objetivo.

La SVR busca encontrar el hiperplano en el espacio de características que tenga la mayor distancia posible a los puntos de entrenamiento que se encuentran dentro de los límites de tolerancia, y que cumpla con el objetivo de minimizar el error. Es útil para problemas de regresión cuando se desea modelar relaciones no lineales y capturar patrones complejos en los datos. Puede manejar datos con características no lineales a través del uso de funciones de kernel, como el kernel polinómico, el kernel gaussiano (RBF) o el kernel sigmoide.

V-C. Árboles de decisión

El árbol de regresión se construye dividiendo recursivamente el conjunto de datos en subconjuntos más pequeños, basándose en las características de los datos. Cada división se realiza seleccionando la característica y el punto de corte que mejor separa los datos en términos de reducir el error de regresión. El objetivo es minimizar la varianza o el error cuadrático medio en cada subconjunto resultante.

A medida que el árbol se construye, las divisiones se realizan de forma recursiva hasta alcanzar un criterio de parada, como alcanzar una profundidad máxima, un número mínimo de

muestras en los nodos o cuando no se obtiene una reducción significativa en el error de regresión.

Una vez construido el árbol, se pueden realizar predicciones para nuevos datos recorriendo el árbol desde la raíz hasta las hojas. Cada hoja contiene un valor estimado de la variable objetivo, que se asigna a la predicción final.

V-D. Bosques aleatorios

En un Bosque Aleatorio para regresión, se construye un conjunto de árboles de regresión independientes. Cada árbol se entrena en un subconjunto aleatorio de datos de entrenamiento y características seleccionados mediante un proceso llamado muestreo bootstrap. Esto implica tomar muestras con reemplazo del conjunto de datos de entrenamiento, lo que crea conjuntos de datos ligeramente diferentes para cada árbol. Además, durante la construcción de cada árbol, en cada división se selecciona un subconjunto aleatorio de características para considerar, lo que introduce más diversidad en el conjunto de árboles.

Durante la predicción, cada árbol de regresión genera una predicción individual y luego se promedian todas las predicciones de los árboles para obtener la predicción final. Este enfoque de promediar las predicciones reduce la varianza y mejora la precisión general del modelo.

VI. ESTRATEGIA DE TRADING

La estrategia planteada busca predecir el precio de cierre de la temporalidad escogida en función del precio de apertura y las demás variables de entrada (ver sección IV). En función de la predicción del modelo, se abre una posición larga/corta dependiendo de si la predicción del precio de cierre del modelo es superior/inferior al precio de apertura.

VI-A. Selección de la temporalidad

La selección de la temporalidad se basa en el estudio de la volatilidad del precio para distintas temporalidades seleccionadas con objetivo de buscar que tanto las comisiones como el diferencial de precio de compra/venta cubran el margen de beneficio obtenido. Para hacer el estudio, se calcula el porcentaje medio de variación por temporalidad durante todo el periodo estudiado, y lo compararé el porcentaje medio de coste fijo por la operativa, para ello se destina un monto inicial para la operativa de 100.000€ con los que realizar las simulaciones.

Las comisiones a la hora de operar en Dukascopy son de 0.1 % por transacción, lo que es equivalente a abrir con una posición porcentual negativa, y dificulta mucho a la hora de tener éxito cuanto menor sea la temporalidad escogida. En el siguiente Cuadro II se muestran las pérdidas fijas debidas a comisiones para la operativa.

Se observa, utilizando Dukascopy Bank como bróker, conforme disminuye la temporalidad la rentabilidad de la operativa disminuye considerablemente debido a las comisiones del bróker. Podemos ver que para la temporalidad diaria, el caso

Temporalidad	Comisiones	Nº Operaciones	Perdida	Volatilidad
10 Minutaria	0,10 %	34312	~ 100 %	0,08 %
30 Minutaria	0,10 %	11555	~ 100 %	0,14 %
Horaria	0,10 %	5887	99,72 %	0,20 %
Diaria	0,10 %	311	26,74 %	0,99 %

Cuadro II: Pérdidas fijas anuales utilizando Dukascopy Bank

más beneficioso, ya partimos de un 26,74 % de pérdida fija, algo que dificulta obtener rentabilidad en el proceso.

Por otro lado, como alternativa se propone Degiro ², un bróker de origen holandés que ofrece una comisión fija de 3€ por transacción, lo que resulta mucho más atractivo para el monto considerado y se adapta mejor a las temporalidades propuestas como se aprecia en el Cuadro III.

Temporalidad	Comisiones	Nº Operaciones	Perdida	Volatilidad
10 Minutaria	6	34312	205,87 %	0,08 %
30 Minutaria	6	11555	69,33 %	0,14 %
Horaria	6	5887	35,32 %	0,20 %
Diaria	6	311	1,87 %	0,99 %

Cuadro III: Pérdidas fijas anuales utilizando Degiro

Utilizando Degiro, reducimos considerablemente las pérdidas fijas de la operativa realizada conforme realizamos menos operaciones para periodos de tiempo más largos. Para el caso diario apenas alcanza el 2 % del monto total, por lo que teniendo en cuenta la suposición de que se aplica un diferencial de precios similar en la operativa de compra y venta de acciones, se selecciona para la simulación de la operativa.

VII. EXPERIMENTACIÓN Y ANÁLISIS DE RESULTADOS

Como se ha comentado anteriormente, el entreno de los modelos y posteriormente, la operativa se realiza sobre la acción de Microsoft.

La implementación de la experimentación se realiza con Python y la librería sklearn. El código de este proyecto está disponible publicamente en el repositorio ³

Primeramente se realiza un estudio de los modelos de predicción haciendo ensayos año a año con todos los datos con los que contamos, y así ver la relevancia en los resultados.

VII-A. Evaluación del modelo de predicción

En este apartado, se entrenan los modelos con apoyo de la librería sklearn y se examinan las métricas para medir la precisión y la capacidad explicativa del modelo en la predicción de valores numéricos.

Los modelos entrenados cubren los mencionados en los alcances y serán denominados de aquí en adelante de la siguiente forma:

RL - Regresión lineal

RP - Regresión polinómica

SVR - Máquina de soporte vectorial

DT - Arbol de decisión

RF - Bosques aleatorios

²<https://www.degiro.es/>

³<https://github.com/mmatar96/automatizacion-de-una-operativa-de-trading-usando-aprendizaje-automatico.git>



Figura 6: Predicciones de los modelos

En el siguiente Cuadro IV se pueden ver los resultados de los ensayos realizados con datos desde el 2017 para los modelos propuestos, destacando el periodo seleccionado inicialmente en los que los modelos obtienen mejores resultados de predicción.

Periodo	RL	RP	SVR	AD	RF
2017/2018	0.96	-4.38	-9.84	-3.55	-3.55
2018/2019	0.99	0.36	-3.65	-1.31	-1.31
2019/2020	-5358	-26.77	-6.27	-2.16	-2.17
2020/2021	0.99	-1.30	-5.29	-2.30	-2.09
2021/2022	0.90	0.48	0.72	0.78	0.76
2022/2023	0.95	0.39	-0.24	0.59	0.71
2020-22/2023	0.99	0.72	0.98	0.96	0.98
2021-22/2023	0.99	0.79	0.97	0.96	0.98

Cuadro IV: Analisis anual 2017/2023 - R cuadrado ajustado

VII-B. Modelo 2021-22/2023

Temporalidad: Diaria (01.01.2021 – 30.08.2023)

Entrenamiento: 01.01.2021 – 31.12.2022

Test: 01.01.2023 – 30.08.2023

En el siguiente Cuadro V se encuentran el resultado del error cuadrático medio, error absoluto medio, error porcentual absoluto medio, R cuadrado y R cuadrado ajustado cuyos acronimos son RMS, MAE, MAPE, R2 y R2_adj respectivamente, para cada uno de los modelos en el periodo seleccionado.

Modelo	RMS	MAE	MAPE	R2	R2_adj
RL	15,48	2,98	1,01 %	0,988	0,988
RP	260,56	11,75	3,85 %	0,806	0,789
SVR	39,45	4,86	1,61 %	0,971	0,968
DT	52,15	5,73	1,93 %	0,961	0,958
RF	51,52	5,73	1,93 %	0,962	0,958

Cuadro V: Modelo 2021-22/2023 - Métricas

Los resultados de la predicción de los modelos se pueden ver en la Figura 6.

VII-C. Indicadores de calidad

Con objetivo de resumir el comportamiento del modelo en la simulación, se muestra en el Cuadro VI algunos indicadores

de calidad para reflejar más detalladamente los resultados del modelo durante la operativa.

Indicador	RL 1	RP	SVR	DT	RF
Exactitud	53 %	52 %	55 %	58 %	57 %
Precisión(+)	55 %	55 %	66 %	59 %	58 %
Sensibilidad(+)	49 %	28 %	22 %	60 %	58 %
F1-Score(+)	52 %	37 %	33 %	59 %	58 %
Precisión(-)	52 %	50 %	52 %	58 %	56 %
Sensibilidad(-)	57 %	76 %	88 %	56 %	56 %
F1-Score(-)	55 %	61 %	66 %	57 %	56 %

Cuadro VI: Indicadores de calidad de la operativa

VII-D. Análisis de rentabilidad

La operativa se simula para los 8 meses del año presente, en los cuales se abrirá y cerrará una posición larga/corta diariamente en función del resultado de la predicción del modelo. Para analizar la rentabilidad se calcula el beneficio obtenido de la simulación.

Se suponen dos tipos de operación: el primer caso haciendo interés simple, en el que se realizan operaciones continuamente con el monto inicial (es decir, 100.000€, para lo que en caso de pérdidas iniciales habría que contar con un capital mayor), y el segundo, haciendo interés compuesto, es decir, se parte del monto inicial y se abren posiciones largas/cortas sumándole el resultado de la operativa, es decir, se reinvierten las ganancias/perdidas).

También se diferencian dos casos, el primero tiene en cuenta una operativa ideal, y el otro incluye las comisiones del bróker seleccionado y un diferencial en el precio entre la oferta y la demanda (el diferencial viene dado por la diferencia en los precios de las ordenes de compra y de venta y se supone un diferencial medio del total de sesiones simuladas), con objetivo de realizar la simulación más real posible.

VII-D1. Resultados de la operativa ideal: En el Cuadro VII encontramos los resultados de la operativa de los modelos en el caso ideal.

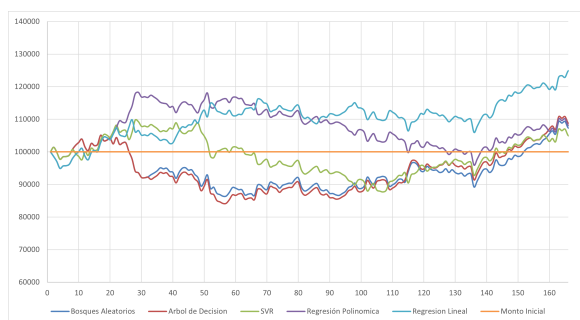
Modelo	Comision	Diferencial	Simple	Compuesto
RL	0	0	24,90 %	26,47 %
RP	0	0	8,10 %	6,89 %
SVR	0	0	8,81 %	7,50 %
DT	0	0	4,99 %	3,52 %
RF	0	0	7,29 %	5,88 %

Cuadro VII: Resultados de la operativa ideal

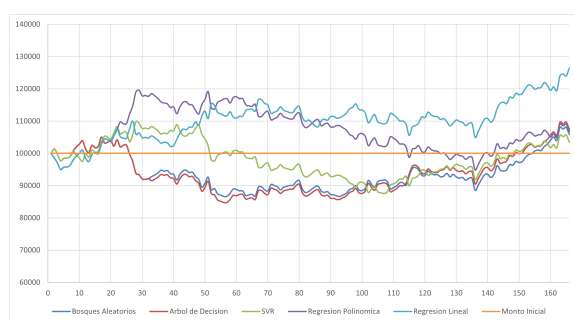
La Figura 7 representa la evolución del capital resultante de aplicar la estrategia planteada por el algoritmo.

VII-D2. Resultados de la operativa real: En el Cuadro VIII encontramos los resultados de la operativa de los modelos en el caso real.

La Figura 8 representa la evolución del capital resultante de aplicar la estrategia planteada por el algoritmo.

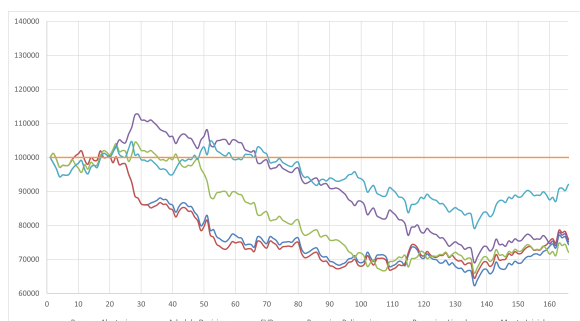


(a)

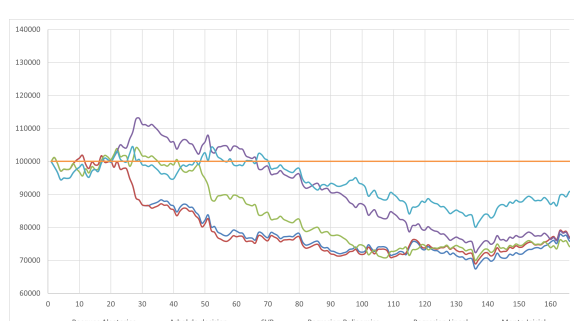


(b)

Figura 7: Evolución del capital en el caso ideal - interés simple (a) y compuesto (b)



(a)



(b)

Figura 8: Evolución del capital en el caso real - interés simple (a) y compuesto (b)

Modelo	Comision	Diferencial	Simple	Compuesto
RL	6	0,193 %	-8,13 %	-9,31 %
RP	6	0,193 %	-24,93 %	-23,42 %
SVR	6	0,193 %	-24,22 %	-23,11 %
DT	6	0,193 %	-28,04 %	-25,92 %
RF	6	0,193 %	-25,74 %	-24,28 %

Cuadro VIII: Resultados de la operativa real

VII-E. Optimización de resultados: Selección de variables

Los resultados obtenidos evaluados para todas las variables de entrada consideradas resulta prometedor, ya que en todas las simulaciones del caso ideal obtenemos un resultado positivo, no tanto en el caso real debido a las pérdidas fijas resultantes de la operativa supuesta. Para finalizar el trabajo, se realiza una serie de ensayos seleccionando distintos subconjuntos de las variables consideradas con objetivo de observar la influencia de cada una de ellas en el modelado de la operativa. Teniendo en cuenta el análisis de correlación, podemos ver que existen algunas variables altamente correlacionadas entre sí, en especial las medias móviles de igual periodo, por lo que son susceptibles a ser evaluadas independientemente ya que puede afectar a la eficacia del modelo. Por otro lado, vemos que las variables RSI, MACD y Volume tienen una correlación relativamente baja, por lo que serán objeto de ser evaluadas individualmente. Finalmente, se destaca la alta correlación entre el precio de apertura y el precio de cierre, lo que hace pensar en una influencia negativa de dicha variable en el modelado. Teniendo en cuenta estas consideraciones, los

subconjuntos de variables a ensayar serán los siguientes:

Ensayo	Variables
1	SMA14, SMA29, SMA50.
2	SMA14_p, SMA29_p, SMA50_p.
3	EMA14, EMA29, EMA50.
4	SMA14, SMA29, SMA50, Vol.
5	SMA14_p, SMA29_p, SMA50_p, Vol.
6	EMA14, EMA29, EMA50, Vol.
7	SMA14, SMA29, SMA50, RSI, MACD.
8	SMA14_p, SMA29_p, SMA50_p, RSI, MACD.
9	EMA14, EMA29, EMA50, RSI, MACD.
10	SMA14, SMA29, SMA50, RSI, MACD, Vol.
11	SMA14_p, SMA29_p, SMA50_p, RSI, MACD, Vol.
12	EMA14, EMA29, EMA50, RSI, MACD, Vol.
13	SMA14, SMA29, SMA50, SMA14_p, SMA29_p, SMA50_p, EMA14, EMA29, EMA50, RSI, MACD, Vol.
14	SMA14, SMA29, SMA50, RSI, MACD, Vol, Open Price.
15	SMA14_p, SMA29_p, SMA50_p, RSI, MACD, Vol, Open Price.
16	EMA14, EMA29, EMA50, RSI, MACD, Vol, Open Price.
17	SMA14, SMA29, SMA50, SMA14_p, SMA29_p, SMA50_p, EMA14, EMA29, EMA50, RSI, Vol.
18	SMA14, SMA29, SMA50, SMA14_p, SMA29_p, SMA50_p, EMA14, EMA29, EMA50, MACD, Vol.

Cuadro IX: Ensayos para selección de variables

Los resultados obtenidos comprenden los promedios de los resultados de todos los modelos entrenados para los casos supuestos y quedan reflejados en el Cuadro X y Cuadro XI.

De los ensayos del 1 al 6 se concluye con que las medias móviles por sí solas, incluyendo o no el volumen, pese a que se ajustan correctamente los datos, no resultan en una operativa eficaz. En los ensayos 7, 8 y 9 reflejan que la inclusión del RSI y el MACD mejora los resultados obtenidos

Ensayo	Exactitud	F1-score	R2
1	52.22 %	57.86 %	0.9209
2	51.98 %	57.59 %	0.9203
3	51.98 %	57.59 %	0.9158
4	51.92 %	59.90 %	0.8236
5	51.98 %	60.00 %	0.8187
6	52.28 %	60.15 %	0.8206
7	53.65 %	57.80 %	0.9768
8	53.53 %	57.64 %	0.9763
9	53.97 %	57.95 %	0.9755
10	57.13 %	60.21 %	0.9004
11	57.13 %	60.09 %	0.9026
12	56.95 %	60.01 %	0.9025
13	54.43 %	57.7 %	0.9119
14	54.49 %	57.58 %	0.9260
15	54.01 %	57.11 %	0.9261
16	54.49 %	57.58 %	0.9252
17	53.71 %	57.02 %	0.9247
18	52.46 %	54.77 %	0.9280

Cuadro X: Analisis del modelo de predicción

Ensayo	Ideal-Simple	Ideal-Comp	Real-Simple	Real-Comp
1	0.83 %	0.28 %	-32.21 %	-28.24 %
2	-0.20 %	-0.98 %	-33.24 %	-29.16 %
3	-0.82 %	-1.65 %	-33.85 %	-29.64 %
4	-1.71 %	-2.92 %	-34.75 %	-30.56 %
5	-2.77 %	-4.01 %	-35.81 %	-31.36 %
6	0.37 %	-0.62 %	-32.66 %	-28.90 %
7	6.31 %	5.22 %	-26.72 %	-24.69 %
8	6.47 %	5.39 %	-26.56 %	-24.57 %
9	6.54 %	5.49 %	-26.50 %	-24.50 %
10	23.95 %	26.55 %	-9.09 %	-9.24 %
11	25.67 %	29.21 %	-7.37 %	-7.32 %
12	24.8 %	27.89 %	-8.23 %	-8.27 %
13	15.17 %	15.02 %	-17.86 %	-17.58 %
14	10.75 %	10.31 %	-22.29 %	-20.98 %
15	7.46 %	6.75 %	-25.58 %	-23.56 %
16	10.31 %	9.94 %	-22.72 %	-21.24 %
17	8.92 %	8.31 %	-24.12 %	-22.45 %
18	5.94 %	5.07 %	-27.09 %	-24.78 %

Cuadro XI: Analisis de rentabilidad

en los ensayos anteriores. En los ensayos 10, 11, 12 y 13, se incluye el Volume para cada una de las medias móviles en los 3 primeros y en el último se tiene en cuenta la misma simulación para el conjunto de todas las medias móviles. En los ensayos 14, 15 y 16 realizamos la misma simulación que en 10, 11 y 12 incluyendo el Open Price. Para terminar, los ensayos 17 y 18 se realizan para todas las variables excluyendo el Open Price en ambos, y el MACD y RSI respectivamente. En estos resultados se puede observar que la inclusión de todas las variables no resulta beneficioso, el Open Price tiene una correlación demasiada alta, lo que influye en contra de la predicción dando como resultado modelos menos eficaces, de igual forma que incluir todas las medias móviles, ya que las de periodos similares tiene una correlación total. Las variables RSI, MACD y Vol parece que aunque no tengan una gran correlación, añadirlas a las variables de entrada mejora drásticamente ya no tanto la predicción, sino operativa.

Con estos resultados se determina que el mejor subconjunto de variables para entrenar el modelo del caso propuesto es el 11: SMA14_p, SMA29_p, SMA50_p, RSI, MACD, Vol.

Finalmente se evalúan los resultados de cada uno de los modelos para el dataset propuesto:

En el siguiente Cuadro XII y XIII se reflejan los cálculos de las métricas de evaluación del modelo de predicción, y los indicadores de calidad respectivamente.

Modelo	RMS	MAE	MAPE	R2	R2_adj
RL	21.57	42.07	14.63 %	0.984	0.983
RP	368.01	13.52	4.42 %	0.727	0.716
SVR	61.37	6.27	2.09 %	0.954	0.953
AD	54.45	5.64	1.93 %	0.960	0.958
RF	75.02	6.76	2.34 %	0.944	0.942

Cuadro XII: Métricas de los modelos de predicción

Indicador	RL 1	RP	SVR	DT	RF
Exactitud	62 %	52 %	52 %	58 %	60 %
Precisión(+)	63 %	53 %	59 %	60 %	62 %
Sensibilidad(+)	64 %	59 %	19 %	53 %	55 %
F1-Score(+)	63 %	56 %	29 %	56 %	58 %
Precisión(-)	62 %	51 %	51 %	57 %	58 %
Sensibilidad(-)	61 %	45 %	87 %	63 %	65 %
F1-Score(-)	61 %	48 %	64 %	60 %	61 %

Cuadro XIII: Indicadores de calidad de la operativa

VII-E1. Resultados de la operativa ideal: El Cuadro XIV refleja los resultados de la operativa en condiciones ideales.

Modelo	Comision	Diferencial	Simple	Compuesto
RL	0	0	34.13 %	38.71 %
RP	0	0	20.73 %	21.39 %
SVR	0	0	-3.86 %	-5.28 %
AD	0	0	42.03 %	50.24 %
RF	0	0	45.82 %	56.08 %

Cuadro XIV: Resultados de la operativa ideal

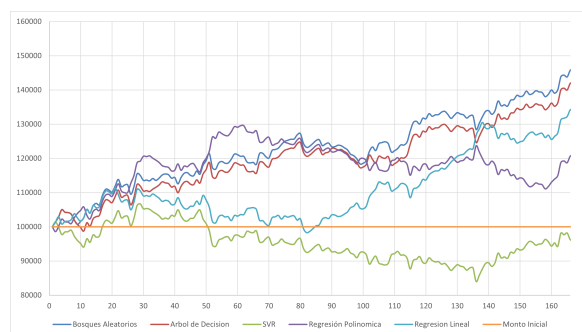
La Figura 9 representa la evolución del capital resultante de aplicar la estrategia planteada por el algoritmo.

VII-E2. Resultados de la operativa real: El Cuadro XV refleja los resultados de la operativa en condiciones reales.

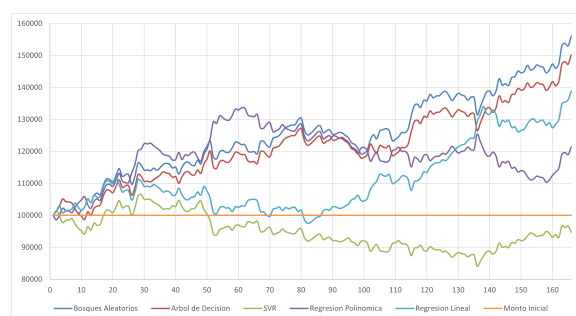
Modelo	Comision	Diferencial	Simple	Compuesto
RL	6	0.193 %	1.09 %	-0.51 %
RP	6	0.193 %	-12.31 %	-12.89 %
SVR	6	0.193 %	-36.89 %	-32.27 %
AD	6	0.193 %	9.00 %	7.89 %
RF	6	0.193 %	12.79 %	12.13 %

Cuadro XV: Resultados de la operativa real

La Figura 10 representa la evolución del capital resultante de aplicar la estrategia planteada por el algoritmo.

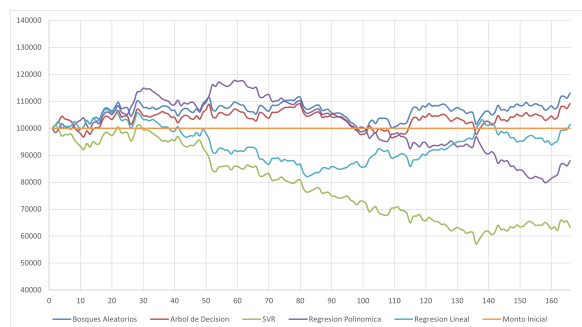


(a)

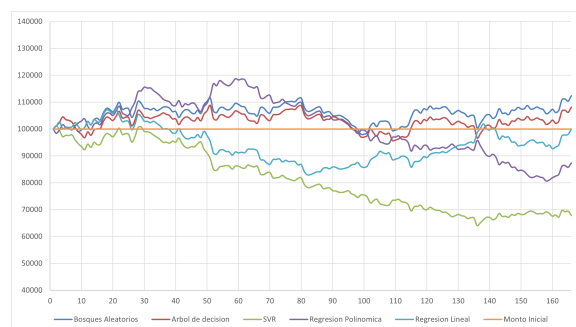


(b)

Figura 9: Evolución del capital en el caso ideal - interés simple (a) y compuesto (b)



(a)



(b)

Figura 10: Evolución del capital en el caso real - interés simple (a) y compuesto (b)

VIII. CONCLUSIONES

Los resultados de los ensayo inicial quedan reflejados en el Cuadro IV y justifican el rango de datos escogido, ya que se comprueba que la similitud entre la distribución de valores de precios es importante a la hora de realizar predicciones. Para el periodo propuesto, casi todos los modelos se ajustan bastante bien a los datos como se aprecia en el Cuadro V, a excepción del modelo de regresión polinómica en el que en general el error de predicción resultante es mayor en comparación con el resto. Los resultados obtenidos en la simulación inicial de la operativa son bastante representativos, ya que pese a que en el caso real supuesto no se consiga obtener una rentabilidad positiva, en el caso ideal todos obtienen beneficio al final de la operativa superando siempre el 50 % de exactitud en la predicción como se aprecia en el Cuadro VI. Esto es debido a las pérdidas fijas por la operativa, por lo que en circunstancias mas beneficiosas de mercado con mayor volumen de negociación, en un broker con mayor nivel de liquidad donde el diferencial de precios de compra/venta sea mas bajo y las comisiones se reduzcan se podria llegar a tener mejores resultados.

Se resalta también en el estudio de la influencia de variables algunas correlaciones peligrosas que afectan negativamente a nuestro modelo y la importancia de seleccionar bien las variables para construir un modelo mas eficiente. Como resultado a la ejecución de los ensayos encontramos un modelo mas eficiente para el caso propuesto. El modelo que mejor se adapta a los datos es la regresión lineal, pero son los modelos

de arbol de decisión y bosques aleatorios los que terminan consiguiendo obtener en la simulación de la operativa mejor resultado ya que aunque la predicción es menos precisa, hace tomar mejores decisiones.

Termino destacando que los resultados obtenidos son muy interesantes por el posible beneficio económico que se puede obtener y existen multitud de posibles enfoques que podrian estudiarse en un futuro utilizando modelos de predicción mas complejos, técnicas de minería de datos para extraer posibles patrones en la variación de precio, incorporar mas variables al estudio como el precio de otros activos con determinada relación o simular diferentes estrategias de trading en función de señales predeterminadas. En definitiva, se concluye este trabajo como una pequeña iniciación a la incorporación de una herramienta en el proceso de toma de decisiones automática que produzca un retorno de inversión significativo y dar la oportunidad de optimizar estrategias de inversión, en la que concluimos con unos resultados positivos y con perspectiva.

REFERENCIAS

- [1] J. J. Elleuch and L. Trabelsi. *Fundamental analysis strategy and the prediction of stock returns*. 2009.
- [2] A. Adebisi, A. Adewumi, and C. Ayo. *Comparison of ARIMA and artificial neural networks models for stock price prediction*. 2014.
- [3] R. Dalio. *Principles*. 2017.
- [4] F. Diebold. *Elements of Forecasting*, 4th ed. 2007.
- [5] F. Fabozzi and P. D.A. *Portfolio Construction, and Analytics*. 2016.
- [6] F. Fabozzi and H. Markowitz. *The Theory and Practice of Investment Management: Asset Allocation, Valuation, Portfolio Construction, and Strategies 2nd*. 2011.

- [7] F. Gers and J. Schmidhuber. *LSTM Recurrent Neural Networks Learn Simple Context Free and Context Sensitive Languages*. 2001.
- [8] D. Lv, S. Yuan, M. Li, and Y. Xiang. *An Empirical Study of Machine Learning Algorithms for Stock Daily Trading Strategy*. 2019.
- [9] J. J. Murphy. *Análisis técnico de los Mercados Financieros*. 1999.
- [10] V.-D. Ta, C.-M. Liu, and D. A. Tadesse. *Portfolio Optimization-Based Stock Prediction Using Long-Short Term Memory Network in Quantitative Trading*. 2020.
- [11] J.-H. Wang and J.-Y. Leu. *Stock market trend prediction using ARIMA-based neural networks*. 1996.
- [12] C. R. y Fernando Sanchez. *Sistemas de aprendizaje automático*. 2022.