TRABAJO FIN DE GRADO

GRADO EN INTELIGENCIA Y ANALÍTICA DE NEGOCIOS

ÁRBOLES DE CLASIFICACIÓN PARA LA TOMA DE DECISIONES SOBRE ACCIONES

ALUMNO/A:	Juan Luis German Saura
TUTOR/A:	Francisco José Goerlich Gisbert
DEPARTAMENTO DEL TUTOR:	Análisis económico
CURSO ACADÉMICO:	2024-2025
FECHA DE DEPÓSITO:	28-05-2025

VNIVERSITAT DE VALÈNCIA Facultat d'Economia

Resumen

El presente trabajo evalúa la capacidad predictiva de modelos de clasificación basados en árboles de decisión (*rpart*) y bosques aleatorios (*randomForest*) para seleccionar acciones del IBEX 35 en función de perfiles de riesgo inversor. A partir de variables financieras transformadas en datos transversales, se construyen reglas de inclusión/exclusión de activos. Los modelos identificados como más idóneos consiguen métricas de rendimiento satisfactorias, siendo la calificación crediticia, la ratio Sharpe y la rentabilidad acumulada los principales determinantes en las decisiones de inversión dependiendo de la clase.

Palabras clave: Árboles de decisión, IBEX 35, clasificación de acciones, perfil de riesgo, inversión

Tabla de contenido

Índice de ilustraciones

1. INTRODUCCIÓN

En un mundo en el que cada vez las decisiones financieras se vuelven más complejas debido al alto grado de dificultad que atañe el entendimiento de los productos financieros, propios de aquellos países más desarrollados, hace que esta tarea cobre constantemente un aumento de importancia para el individuo, debido a que mayores grados de libertad suponen una mayor autonomía en la elección de las decisiones propias, pero con la contrapartida de una mayor responsabilidad personal sobre la opciones que acaban siendo escogidas[1].

Sumado al aumento del desarrollo tecnológico a escala global, la accesibilidad a la adquisición de activos financieros ha ido también en aumento, por lo que existe una mayor proporción de personas que han adquirido posesión sobre activos a través de este método[2], impulsando el nacimiento de nuevos inversores minoristas está en auge. Estas tendencias se han ido aconteciendo igualmente en elementos altamente relacionados entre sí como lo son los fondos indexados, ETFS y acciones[3].

A causa de lo anterior, la toma de decisiones en la adquisición de activos financieros se ha vuelto cada vez más compleja, debido a la expansión de los mercados. Discernir en entornos en los que prima la volatilidad y la incertidumbre, en los que el riesgo siempre está presente y añadiendo que existe una altísima cantidad de información para analizar[4], todo ello crea la necesidad de encontrar herramientas que ofrezcan soluciones relativamente simples a la problemática planteada.

De esta forma, la razón principal por la cual se ha realizado el trabajo, es dar respuesta a este tipo de decisiones tan vitales para el contexto financiero del creciente número de inversores particulares que cada vez se interesan más por mantener igual o aumentar su bienestar financiero[5], que influye directamente en las perspectivas vitales del sujeto. Consecuentemente, se propone como hipótesis principal determinar si los árboles de decisión son una herramienta efectiva para clasificar acciones del IBEX 35 en función de variables clave y diferentes perfiles de riesgo.

Por consiguiente, la aplicación de modelos de *machine learning* aplicados a los problemas de clasificación, como son los árboles de decisión/clasificación[6], son una de las varias herramientas existentes que pueden ser aplicadas para la resolución de estas situaciones desarrolladas bajo panoramas sofisticados.

Gracias a estos algoritmos, se dispone de una identificación sencilla sobre las reglas de clasificación que se utilizaron para la inclusión o no sobre los elementos analizados en una cartera de activos. Las reglas se crearán a partir de las variables consideradas que tratan circunstancias claves en la exploración de los activos, como son la volatilidad, la rentabilidad histórica, la calificación crediticia, etc.

Debido a esto, la aplicación de los patrones encontrados puede ayudar a comprobar que factores son determinantes en la toma de decisiones, simplificando ampliamente el proceso de análisis, brindando soluciones abreviadas para los inversores particulares sobre las variables más significativas, dependiendo del nivel de riesgo que toleraría cada inversionista.

En cuanto a la revisión de la literatura, sobre investigación de temas similares, los autores Yehan Wang (2024)[7] utilizaron árboles de decisión con variables parecidas, pero el objetivo principal era la predicción del precio de la acción junto con maximizar la rentabilidad de la cartera, sin tener en cuenta los perfiles de riesgo. Por otra parte, Klokholm y Thomsen (2025)[8], vuelve a utilizar factores similares para la predicción, pero con el fin de optimizar la composición de la cartera de activos.

Para llevar a cabo este cometido, se han recopilado, procesado y manipulado los datos mensuales entre los años 2016-2024 de las empresas con más ponderación pertenecientes al índice del IBEX 35. De esta manera, quedaría evaluar la calidad de las predicciones realizadas por los modelos según la disposición al riesgo de cada persona, que se organizaría en tres niveles según su aversión al mismo, los cuales serían: el perfil tolerante, el conservador y el neutro.

Con ello, se establecería como objetivo principal del trabajo el reconocimiento de las variables clave para la inclusión de activos en un portfolio para mejorar la toma de decisiones financieras sobre esta clase de activos.

Otros fines secundarios que se pretende tratar de alcanzar serían los siguientes: tratar de agilizar el análisis de los grandes volúmenes de datos para sintetizar y examinar los rendimientos de los modelos entrenados según el nivel de riesgo para cada inversor particular.

De este modo, para la consecución de todo lo que se ha ido exponiendo en la introducción, el resto del trabajo se estructurará en un cuerpo central, mediante la determinación de un marco teórico donde se situarán las bases teóricas y las hipótesis

a investigar, la metodología en la que se explicarán los instrumentos utilizados, los resultados con su pertinente análisis junto con su explicación y con una comparativa con el marco teórico previo, finalizando con las conclusiones presentadas como síntesis del trabajo.

2. MARCO TEÓRICO

2.1 Árboles de decisión

Se definen los árboles de decisión como algoritmos de *machine learning* que permiten la construcción de modelos predictivos capaces de ser aplicados a problemas de índole clasificatoria o relacionados con la regresión, lo que resalta el gran abanico de posibilidades a desarrollar y la aplicabilidad que pueden llegar a tener. Su valor radica en su capacidad para analizar grandes volúmenes de datos (*Big Data*), identificar patrones y tomar decisiones basadas en la información obtenida[9].

Sobre la aplicación en el campo de la regresión, el fin central es la predicción de una variable continua a partir del resto de variables independientes encontradas en el conjunto de datos utilizado. Como ejemplo del rendimiento que es posible obtener, los autores Máximo Camacho, Salvador Ramallo, Manuel Ruiz Marín (2021) [10], determinaron en su estudio del precio de la vivienda, la habilidad que atesoran este tipo de modelos.

En cuanto a los modelos de clasificación, estos utilizan variables binarias para categorizar los datos. Para este caso específico, esta variable binaria es manipulada según umbrales definidos sobre cada factor para cada perfil inversionista, por lo que está variable categórica puede cambiar en su valoración “incluido” o “no incluido” para cada empresa según cada escenario[11].

Los árboles de clasificación constan en su formación de nodos y ramas y siguen una estructura jerárquica propia de un árbol como el mismo nombre indica. El árbol tiene en la parte superior el nodo raíz, que representa la condición a evaluar sobre la que se centra el árbol. En el siguiente nivel estarían los nodos internos, que representan las pruebas lógicas (reglas) intermedias, que se dividen en función de las distintas características que considera el árbol. En el nivel más bajo están los nodos hoja, que son los puntos terminales del árbol donde se llega a una decisión concreta y no tienen más ramas salientes. Todos estos nodos se encuentran por unidos por las ramas, que

son las encargadas de conectar los nodos según los resultados de las pruebas lógicas presentadas en los puntos intermedios de los nodos internos[12].

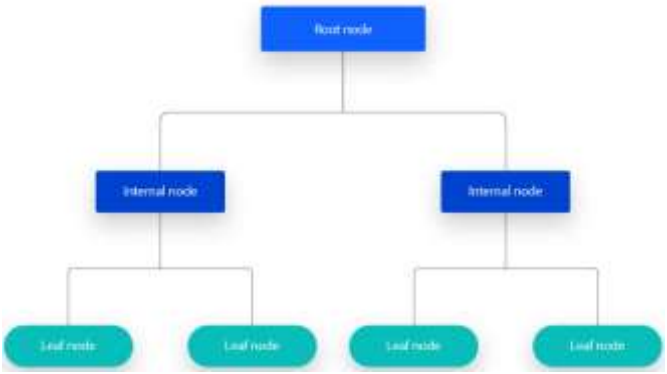

Ilustración 1: Ejemplo árbol de decisión

Estas herramientas, además de las comentadas en la introducción, utilizan árboles de decisión mediante factores como la calificación crediticia. Con estas bases como antecedentes, se encuentra suficiente evidencia para ejemplificar como sirven para simplificar la toma de decisiones en la selección de activos dentro del ámbito financiero.

Por ello, como ventajas de los árboles de decisión es fácilmente visible la gran interpretabilidad que proporcionan, ya que exponen de manera clara los patrones encontrados de forma totalmente visual, por lo que las características de justificabilidad y transparencia están más que presentes en los árboles de clasificación. Asimismo, permite manejar variables tanto categóricas como numéricas, lo que permite una mayor flexibilidad a la hora de aplicar este tipo de modelos a conjuntos de datos más diversos, causando que se puedan adaptar a distintos contextos.

Sin embargo, el punto negativo que se debe tener en cuenta sobre los árboles de clasificación es la gran propensión que tienen en caer en errores de sobreajuste, ajustándose en exceso a los datos de entrenamiento, ello llevando a la creación de modelos poco robustos que sufren de variaciones altamente significativas en su rendimiento y actuación con pequeñas modificaciones en el conjunto de datos. Para ello existen métodos de poda o bien *random forest* para mitigar esta debilidad. Relacionado con ello, los árboles de decisión pueden sufrir en la precisión de las predicciones que generarían si existiera un desbalance de clases.

2.2 Variables clave y perfiles de riesgo

La tarea de escoger entre una vasta gama de opciones entre los activos financieros adquiribles, como son las acciones, depende de las deducciones que se hagan sobre las variables que miden diferentes aspectos. La importancia que tienen se debe a la oportunidad que estas representan para los inversores en cuanto a la toma de decisiones informadas, que a través del análisis de las métricas, permite el mayor ajuste posible a las necesidades de cada inversionista en la distinción de cada instrumento financiero[13].

De esta forma, los elementos cruciales que han sido capaces de ser recopilados desde fuentes públicas serían aquellos que examinan la valoración de la acción, el riesgo, la rentabilidad pura y la rentabilidad ajustada al riesgo. Mediante estas categorías, es cuantiosamente más sencillo entender que dimensiones son las más significativas para la tarea de la clasificación que realiza el árbol de decisión.

Por medio de esta agrupación, se puede establecer que, a rasgos generales, para los inversores que caigan en los distintos grupos diferenciados según el factor de la inestabilidad, los inversores con alta aversión al riesgo (conservadores), se inclinarán hacia posturas donde la volatilidad global de la acción sea baja y la rentabilidad de la misma sea mínimamente positiva. Para aquellos con una condición de baja aversión al riesgo (tolerantes), estos pueden llegar a permitirse correr mayores riesgos para tratar de alcanzar las mayores rentabilidades posibles. En cuanto a los tipos intermedios en la aversión, estos buscarían un equilibrio entre ambas dimensiones estudiadas.

De igual modo, se debe resaltar que estas variables presentan algunas limitaciones. Por un lado, la imposición de solo poder explorar a partir del año 2016 reduce bastante la amplitud de la capacidad de recogida de datos, ya que hay empresas que tienen información disponible desde el año 2000 que no se puede aprovechar al tener una obligación en mantener el mismo horizonte temporal para todas las compañías consideradas.

Además, los cambios en la totalidad del mercado pueden afectar a selectas variables al depender en sus cálculos directamente de este, por lo que circunstancias no tomadas en cuenta en el análisis pueden seguir teniendo un impacto relevante, por lo que no existe un escenario totalmente independiente ni manipulable para los fines del trabajo, ya sea empresas pertenecientes al índice del IBEX 35 u otros mercados internacionales.

3. DATOS Y FUENTES

3.1 Descripción del IBEX 35 y selección de empresas

Un índice bursátil es un indicador histórico que recoge todas las acciones para cada empresa que compone dicho índice. De esta manera, un índice bursátil sirve a un propósito de simplificación de información, ofreciendo la evolución de los valores de las acciones de todas las acciones que tiene en cuenta[14].

Por esto, el índice contemplado para este proyecto ha sido el más importante dentro del mercado financiero español, que es el IBEX 35. La transcendencia de este índice surge porque se encarga de registrar a las empresas con mayor volumen de negociación dentro del mercado español, que a su vez reúne las bolsas de Madrid, Barcelona, Bilbao y Valencia. Se debe mencionar, que la composición de este índice es variable, ya que pueden existir movimientos de entrada y de salidas para aquellas empresas que aumentan o disminuyen su liquidez, por lo que la constitución del IBEX 35 está en constante revisión por un Comité Asesor Técnico (CAT)[15].

Así pues, para llevar a cabo el cometido las intenciones diseñadas, se ha optado por realizar una criba sobre el índice para reunir los datos sobre las 20 empresas más simbólicas de todos los sectores presentes, priorizando aquellas que más peso y ponderación tienen según la capitalización bursátil dentro del índice (anexo 1)[16]. Esto se ha realizado con la intención de encontrar un equilibrio entre representatividad de la diversidad del índice y la simplicidad, es decir, los modelos generados buscan minimizar el sobreajuste al equilibrar la estabilidad del análisis con la naturaleza dinámica del mercado financiero español[17].

Por tanto, se ha tratado evitar el sesgo de muestreo con este enfoque de abundante variedad sectorial de las compañías tomadas en cuenta, para así conseguir una generalización robusta sacrificando cierta especificidad por falta de cobertura total del mercado financiero, es decir, la problemática de la compensación entre el sesgo y varianza[18].

Entrando en más detalles de las empresas por las que se ha optado en el análisis, el primero de todos los sectores consideraros ha sido el financiero, en el que están bancos como: *Santander, Caixabank, BBVA, Sabadell* y *Bankinter*, encargados de ofrecer servicios financieros. Este tipo de actividad es uno de los pilares más

importantes del IBEX 35, al pertenecer un gran número de este tipo de compañías a este mismo sector y ser a su vez empresas internacionalizadas algunas de ellas.

El segundo sector a considerar es el energético, en el que han entrado: *Iberdrola, Endesa, Repsol*, *Redeial* y *Naturgy* encargadas de una serie de tareas como la distribución de la electricidad, el gas y el petróleo.

El tercer ámbito empresarial es el referente a la telecomunicación y la tecnología, en el que se encuentra a la reconocida *Telefónica* (presencia en Latinoamérica), junto con *Amadeus, Cellnex* y *Acciona.*

El antepenúltimo sector ha sido el relacionado con la industria y construcción, donde localizan *Ferrovial* y *ACS*, delegadas a ofrecer respuesta al mercado inmobiliario.

El penúltimo sector es el de consumo, donde se agrupan *Inditex*, *IAG* y *Aena*. Aunque las actividades de estos negocios difieran entre sí, todos tienen en común la venta al por menor.

Por último, el sector de la salud lo representa la empresa *Grifols*, centrándose en todo lo relacionado con la producción y venta del plasma a nivel internacional[19].

Esta diversificación sectorial permite conseguir y seguir la aproximación establecida en párrafos anteriores, por lo que se trata de obtener la imagen más amplia del principal mercado financiero español sin tener que reunir la información de todas las empresas integrantes del índice bursátil.

3.2. Precios de cierre históricos

Una vez ya establecidas las empresas sobre las que se va a realizar el trabajo de recolección de datos, la fuente más asequible por su naturaleza, al ser de libre acceso y gratis, ha sido la denominada como *Yahoo Finance,* una plataforma diseñada para proporcionar datos financieros, reportes financieros, noticias de prensa relacionadas con el ámbito y almacenar públicamente los registros históricos de compañías con presencia en bolsa.

Esta base de datos pública es de extrema utilidad porque permite la posibilidad de extraer la variable base a partir de la cual se crean el resto de las variables derivadas, que es el precio de cierre de la acción. Como ejemplo de esto, algunos de los indicadores considerados en la creación de los modelos, como las medidas de riesgo (volatilidad[20] y beta), de rentabilidad[21] (acumulada, mensual y relativa) y de rentabilidad

ajustada al riesgo (alfa de Jensen[22] y ratio de Sharpe[23]) se calculan directamente con el precio de cierre de la acción.

El precio de cierre es la muestra del último precio en el que se cerró la última transacción de un activo concreto durante el horizonte temporal escogido. De esta manera, el precio de cierre se convierte en un elemento central para cualquier persona interesada en el campo de la bolsa, ya que se transforma en la referencia base al reflejar el comportamiento y la valoración por parte de los inversores de un activo determinado. Gracias a esto, si se analiza el precio de cierre histórico de cualquier activo, es posible captar las tendencias generales sobre el comportamiento y valoración de dicho activo[24].

Sumado a esto, *Yahoo Finance* ofrece más ventajas aparte de la posibilidad de obtener el precio de cierre a través de sus datos. Estos puntos positivos son la sencilla integración que ofrece con el lenguaje de programación que se utiliza para el tratamiento de los datos (R), mediante el paquete *quantmod*[25], que facilita la automatización del análisis. Además, la plataforma ofrece una cobertura histórica lo suficientemente amplia como para llevar a cabo el análisis con una cantidad de información aceptable para la ejecución del proyecto.

Teniendo esto presente, la frecuencia temporal por la que se ha optado obtener de entre las disponibles (anual, semestral, trimestral, mensual y diaria) ha sido la mensual, fundamentalmente por el hecho de que se elimina la posibilidad de tener en los datos descargados valores faltantes, que dificulten las tareas posteriores. Añadido a esto, la frecuencia mensual permite poder evaluar tendencias más a largo plazo que si se escogiera, por ejemplo, una frecuencia diaria.

No obstante, la única limitación con respecto a los precios de cierre en bolsa, es que no todas las empresas coinciden en el mismo horizonte temporal, ya que la idea inicial fue examinar el período comprendido entre los años 2000-2024 y eventualmente se ha tenido que reducir este lapso significativamente hasta los años 2016-2024, a fin de garantizar la coherencia y calidad de los datos, que es donde todas las empresas coincidían sobre el registro de datos histórico y así poder realizar tareas de comparación entre las métricas generadas, mediante un conjunto de datos homogéneo y consistente.

Por estas razones, los datos que ofrece *Yahoo Finance* permiten analizar los datos de los precios de cierre históricos, pero su análisis detallado escapa al alcance de este estudio.) y las variables relacionadas, lo que concede la opción de tener la base de

datos necesaria para partir hacia el entrenamiento de los modelos de clasificación, evaluando la idoneidad de cada activo en función de sus características financieras.