



Facultad de Ciencias Económicas y Empresariales

# **INTELIGENCIA ARTIFICIAL COMO HERRAMIENTA PARA LA INVERSIÓN EN EL MERCADO DE VALORES**

Marina Bellvert Torres  
Clave: 201503073

MADRID | abril 2020

Marina Bellvert Torres

**INTELIGENCIA ARTIFICIAL COMO HERRAMIENTA PARA LA INVERSIÓN EN EL  
MERCADO DE VALORES**



## Título: Inteligencia Artificial como herramienta para la inversión en el mercado de valores

### Resumen:

Uno de los pilares fundamentales dentro del mundo de la inversión se basa en la obtención del mayor beneficio posible con las operaciones realizadas, lo que ha llevado al desarrollo de numerosos métodos y herramientas enfocados a lograr tal fin. A pesar de la gran variedad de métodos y percepciones que han ido surgiendo a lo largo de los años, uno de los que más interés ha suscitado históricamente ha sido la predicción de los movimientos del mercado.

En este sentido, muchos han sido los investigadores interesados en encontrar las herramientas idóneas para la obtención de predicciones más precisas, pues estas permitirían lograr un beneficio mayor. A pesar de tratarse de un ámbito ampliamente estudiado, el desarrollo de tecnologías cada vez más avanzadas y el surgimiento de la inteligencia artificial, han abierto la oportunidad de continuar investigando en esta materia.

Por ello, el objetivo principal de este trabajo se centrará en investigar cómo se está utilizando la inteligencia artificial como herramienta para la inversión en el mercado de valores y si su uso está relacionado con posibles mejores resultados. Gracias al análisis de la literatura y de los estudios empíricos disponibles, nos será posible comprobar la interesante oportunidad que plantea el uso de inteligencia artificial para el inversor. Por otro lado, también se descubrirán cuáles son los métodos utilizados con este fin y cuáles son los aspectos que deben de tenerse en cuenta para utilizar estas herramientas de forma correcta.

Palabras Clave: Inversión, inteligencia artificial, mercado de valores, *machine learning*, modelos, algoritmos.

### Abstract:

One of the cornerstones of the investment business is the pursuit of the highest yield, which has led to the development of several methods and tools. Despite the great variety

of methods and perceptions that have emerged over the years, stock market prediction has been one of the most widely analyzed.

In this regard, many researchers have been interested in finding the right tools for obtaining more accurate predictions, as these would allow for greater benefit. Despite this has been a widely studied area, the development of advanced technologies and the emergence of artificial intelligence has opened up the opportunity to continue investigating this field.

For this reason, the main objective of this project will be investigating how artificial intelligence is being used as a tool for investment in the stock market and whether its use is related to better results. Thanks to the analysis of the literature and the available empirical studies, we will be able to clarify the interesting opportunity that the use of artificial intelligence poses for the investor. On the other hand, we will also analyze which are the methods used for this purpose and which are the aspects that must be taken into account in order to use these tools properly.

Key words: Investment, artificial intelligence, stock market, *machine learning*, models, algorithms.

# TABLA DE CONTENIDOS

<b>1.</b>	<b>INTRODUCCIÓN</b> .....	<b>7</b>
1.1.	PROPÓSITO PRINCIPAL Y OBJETIVOS .....	9
1.2.	MOTIVACIÓN .....	9
1.3.	METODOLOGÍA .....	10
1.4.	ESTRUCTURA .....	11
<b>2.</b>	<b>INTRODUCCIÓN A LA INVERSIÓN Y EL TRADING</b> .....	<b>12</b>
<b>3.</b>	<b>PREDICCIÓN DEL MERCADO DE VALORES</b> .....	<b>15</b>
3.1.	HIPÓTESIS DEL MERCADO EFICIENTE.....	15
3.2.	HIPÓTESIS DEL PASEO ALEATORIO .....	16
3.3.	INTRODUCCIÓN A LA PREDICCIÓN DEL MERCADO DE VALORES .....	17
3.4.	MODELOS ESTADÍSTICOS PARA LA PREDICCIÓN DEL MERCADO DE VALORES .....	19
<b>4.</b>	<b>INTELIGENCIA ARTIFICIAL EN LA PREDICCIÓN DEL MERCADO DE VALORES</b> .....	<b>20</b>
4.1.	INTRODUCCIÓN A LA INTELIGENCIA ARTIFICIAL .....	20
4.2.	APLICACIÓN DE LA INTELIGENCIA ARTIFICIAL AL MERCADO DE VALORES.....	25
4.2.1.	ARTIFICIAL NEURAL NETWORKS .....	27
4.2.1.	SUPPORT VECTOR MACHINES .....	29
4.2.1.	RANDOM FOREST .....	31
<b>5.</b>	<b>ANÁLISIS DE ESTUDIOS</b> .....	<b>33</b>
5.1.	ARTIFICIAL NEURAL NETWORKS.....	34
	<i>Estudio 1 – Pakdaman Naeini, Taremian y Baradaran Hashemi (2010)</i> .....	34
	<i>Estudio 2 – M. Harper (2016)</i> .....	36
	<i>Estudio 3 – Qiu y Song (2016)</i> .....	38
5.2.	SUPPORT VECTOR MACHINES .....	42
	<i>Estudio 4 – Saahil Madge (2015)</i> .....	42
	<i>Estudio 5 – Di (2014)</i> .....	43
	<i>Estudio 6 – Yu, Wang y Keung Lai (2005)</i> .....	45
5.3.	RANDOM FOREST .....	47
	<i>Estudio 7 - Khaidem, Saha y Roy Dey (2016)</i> .....	47
	<i>Estudio 8 - (Bafandeh Inmandoust &amp; Bolandraftar, 2014)</i> .....	50
	<i>Estudio 9 – Basak, Saha, Kar, Khaidem y Dey (2019)</i> .....	52
<b>6.</b>	<b>CONCLUSIONES</b> .....	<b>54</b>
<b>7.</b>	<b>BIBLIOGRAFÍA</b> .....	<b>57</b>
<b>8.</b>	<b>ANEXO</b> .....	<b>62</b>

## Índice de Tablas

Tabla 1. Clases de indicadores técnicos y ejemplos.....	14
Tabla 2. Resumen estudios basado en la predicción mercado de valores con IA .....	33
Tabla 3. Modelo ANN 1 .....	35
Tabla 4. Modelo ANN 2 .....	38
Tabla 5. Comparación modelos. Estudio ANN 3 .....	39
Tabla 6. Modelo ANN 3 .....	40
Tabla 7. Indicadores técnicos tipo 1. Modelo ANN3 .....	41
Tabla 8. Indicadores técnicos tipo 2. Modelo ANN 3 .....	41
Tabla 9. Modelo SVM 1 .....	43
Tabla 10. Comparativa de la precisión. Modelo SVM 2 .....	44
Tabla 11. Modelo SVM 2 .....	44
Tabla 12. Comparación modelos. Modelo SVM 3 .....	46
Tabla 13. Modelo SVM 3 .....	46
Tabla 14. Resultados para las acciones de Apple Inc. Modelo RF 1 .....	48
Tabla 15. Comparación modelos. Modelo RF 1 .....	49
Tabla 16. Modelo RF 1 .....	49
Tabla 17. Indicadores Técnicos. Modelo RF 1 .....	50
Tabla 18. Indicadores técnicos modelo RF 2. ....	51
Tabla 19. Modelo RF 2 .....	52
Tabla 20. Modelo RF 3 .....	54

## Índice de Ilustraciones

Ilustración 1. Categorías incluidas en la predicción de precios.....	18
Ilustración 2. Simplificación de un modelo.....	22
Ilustración 3. Representación gráfica de la capacidad de generalización de un modelo de Machine Learning.....	24
Ilustración 4. Representación SVM.....	30
Ilustración 5. Simplificación modelo Random Forest.....	31
Ilustración 6. Curva ROC Apple Inc. Modelo RF 1 .....	48

## 1. INTRODUCCIÓN

Las sociedades actuales se encuentran sumidas en un gran y revolucionario proceso de cambio tecnológico. Asimismo, la velocidad de los cambios a los que nos enfrentamos requiere de una adaptación constante por parte de la sociedad. En este contexto, empresas y particulares a nivel internacional dedican grandes esfuerzos para tratar de sacar el máximo beneficio posible de esta nueva tendencia, invirtiendo grandes cantidades de dinero en tecnologías de la información y buscando la forma de adaptar sus procesos productivos a estas nuevas posibilidades.

El avance tecnológico trae consigo el desarrollo de herramientas capaces de procesar grandes cantidades de información a una velocidad antes inimaginable, superando con creces las posibilidades humanas en este aspecto. Además, no únicamente se han desarrollado capacidades de análisis, sino que estas mismas máquinas son capaces de ofrecer resultados en base a la información analizada mediante el aprendizaje.

Gana así importancia el término *modelo*, que puede entenderse como una versión abstracto-matemática que trata de reflejar la realidad para poder analizarla de una forma más sencilla. A pesar de que los modelos llevan con nosotros muchos años el avance tecnológico, al que se hacía alusión previamente, ha permitido su desarrollo y evolución, llevándonos a modelos más complejos, capaces de procesar más información y de elaborar conclusiones mucho más precisas. Por este motivo, surge un gran interés por parte de las empresas de nuestro entorno en incorporar modelos a su actividad diaria con el fin de que estos evalúen la información disponible y tomen decisiones de una forma más rápida y precisa.

Esta situación ha convertido la utilización de modelos en una de las grandes tendencias de nuestra sociedad; la constante necesidad de adaptación para ser competitivo en el mercado lleva a una incesante búsqueda de la eficiencia y la reducción de costes, motivo por el cual todo proceso susceptible de ser automatizado o mejorado mediante modelos suscita un gran interés.

El mundo de la inversión no ha quedado exento de esta nueva tendencia; la capacidad actual de las máquinas de analizar información supera con creces la capacidad humana, y esto también ha visto su potencial aplicación en el ámbito los activos cotizados. El

objetivo principal de las operaciones realizadas en el mercado de valores reside en la obtención del máximo beneficio posible, por lo que gestores de todo el mundo utilizan los recursos de los que disponen con este fin. En este sentido, la aplicación del desarrollo tecnológico ha permitido la ampliación de estos recursos, permitiendo la incorporación de nuevas herramientas que podrían ser de gran utilidad a la hora de obtener una rentabilidad mayor.

Aparecen así diferentes ramas dentro de la inversión que buscan aprovecharse de este tipo de herramientas. Una de estas ramas se centra en la utilización de la inteligencia artificial, y más concretamente, *machine learning*, para tratar de predecir los movimientos del mercado con el fin de obtener una rentabilidad mayor. A pesar de que durante muchos años la posibilidad de predecir los movimientos del mercado ha sido muy criticada e incluso rechazada, la aparición de estas nuevas herramientas abre una nueva oportunidad para muchos inversores.

Es por ello por lo que cada vez son más los gestores y bancos que apuestan por la utilización de modelos que emplean la inteligencia artificial para predecir los movimientos del mercado, al igual que cada vez son más los expertos en *machine learning* que han centrado su actividad en este ámbito.

Sin embargo, a pesar de lo interesante de estos instrumentos, todavía su utilización en ciertos países, como es el caso de España, es significativamente baja. Este hecho guarda importante relación con el desconocimiento generalizado tanto de la industria financiera como tecnológica, lo que lleva incluso a producir rechazo entre la población. Igualmente, más allá del público general, muchos son los gestores que reniegan de utilizar este tipo de herramientas por desconocimiento del mundo digital y otros muchos son los expertos en inteligencia artificial que, por desconocimiento de su posible aplicación al mundo financiero, no deciden emplear sus conocimientos en dicho sector.

Por este motivo, surge un interés generalizado en acercar ambos conocimientos para poder extraer todo el potencial que tienen y poder así continuar avanzando en esta materia.



## 1.1. Propósito principal y objetivos

El propósito general de este trabajo se centra en mostrar cómo la inteligencia artificial puede ser una herramienta útil para obtener beneficios en el mundo de la inversión y el trading mediante la predicción de los movimientos del mercado; la pregunta de investigación que nos interesa responder con este trabajo se basa en si es útil el uso de inteligencia artificial para elaborar una estrategia de inversión. Por lo tanto, la hipótesis que se tratará de verificar es la siguiente: *“La inteligencia artificial puede servir como herramienta útil para la inversión en activos cotizados”*.

Junto con el propósito principal, el presente trabajo busca responder a otros objetivos secundarios que se plantearán a continuación.

A lo largo de esta tesis se presentarán teorías ampliamente extendidas que afirman la imposibilidad de predecir los movimientos del mercado de valores, uno de los objetivos secundarios de nuestro trabajo será tratar de refutar dichas teorías. Por otro lado, se buscará hacer un análisis de los modelos utilizados para la predicción del mercado de valores, tratando de encontrar el modelo más adecuado para este fin. De igual manera, este trabajo tiene la intención de servir de punto de unión entre el mundo de la inversión y el mundo de la inteligencia artificial, ayudando a establecer una base sobre la que se podrán desarrollar investigaciones futuras.

## 1.2. Motivación

Nuestras circunstancias actuales están fuertemente determinadas por el constante proceso de desarrollo tecnológico al que nos enfrentamos; cada vez son más los sectores, profesiones y sociedades afectadas por la revolución digital, y esta tendencia no ha hecho más que comenzar. Por otro lado, la rapidez con la que este proceso está teniendo lugar hace que sea necesaria la investigación constante con el fin de sacar el máximo partido de esta situación.

En este contexto, el rápido avance de la inteligencia artificial y su aplicación a diversas áreas han abierto un interesante debate sobre la posibilidad de hacer de esta una útil herramienta para la inversión. Sin embargo, el desconocimiento generalizado y la poca concreción de la información disponible dificulta mucho la extensión de su uso por parte del ciudadano, incluso cuando este tiene conocimientos financieros. En este aspecto,

mientras que en otros países el uso de inteligencia artificial en la inversión está muy extendido, actualmente en España el conocimiento y trabajo con este tipo de herramientas sigue en plena expansión (Seth, 2019), por lo que debe incentivarse su desarrollo de cara a la mejora de la competitividad internacional.

Nos encontramos por tanto en un momento que requiere de nuestra implicación y esfuerzo para poder extraer las ventajas que estas nuevas tecnologías nos brindan. Por este motivo, creemos necesario fomentar el acercamiento del mundo tecnológico al financiero de forma que sea posible el aumento del interés de la cuestión.

Con esta finalidad, a lo largo del presente trabajo se explicarán cuáles son las herramientas de inteligencia artificial de las que disponemos hoy en día y cómo se utilizan para optimizar la inversión con un objetivo doble: por un lado, probar lo interesante del uso de la inteligencia artificial en la inversión, y, por otro lado, servir de puente de unión entre el mundo tecnológico y el financiero.

### 1.3. Metodología

Para llevar a cabo la elaboración del presente trabajo y la consecución del propósito principal y los objetivos de este se empleará una metodología deductiva, lo quiere decir que se tratará primero lo más general y se irá entrando en lo más específico a medida que transcurre el trabajo.

En primer lugar, se procederá a exponer qué es el mercado de valores y con qué objetivo se buscaría hacer predicciones sobre este, profundizando en los conceptos básicos para la comprensión posterior de la investigación.

Posteriormente, se procederá a explicar la literatura referente a la predicción del mercado de valores y su evolución, para finalmente exponer cómo la inteligencia artificial puede utilizarse en este ámbito. Previamente a la introducción de la inteligencia artificial como herramienta de inversión, se explicará qué es *machine learning* y cómo funciona, así como diversos términos relacionados con este tipo de tecnología, tratando de facilitar nuevamente el entendimiento del lector.

Una vez se haya introducido cómo la inteligencia artificial se relaciona con el mercado de valores, se pasará a analizar estudios concretos en los que pueda verse la aplicación práctica de estas herramientas.

En último lugar, con el fin de responder a la pregunta de investigación, se presentarán las conclusiones extraídas del análisis elaborado.

En lo referente al tipo de información empleada en este trabajo, se ha recurrido a publicaciones académicas y libros para la presentación teórica, así como a estudios empíricos realizados por diversos investigadores para mostrar la exposición práctica. Tanto los estudios como las publicaciones académicas han sido extraídas de bases de datos como Jstor, Scopus, Google Scholar o Research Gate.

#### 1.4. Estructura

La estructura del presente trabajo se divide en cuatro partes. En el primer bloque (apartado 2) se introduce el concepto “mercado de valores” y se explica por qué puede darse un interés en predecir sus movimientos. En este apartado también se explican algunos de los conceptos técnicos más utilizados a lo largo de la investigación.

El segundo bloque (apartados 3 y 4) se centra en el estudio de la predicción del mercado de valores, analizando las teorías surgidas en torno a este ámbito e investigando acerca de los métodos que se han utilizado con esta finalidad. Este apartado finalizará con la incorporación de la inteligencia artificial como herramienta de predicción y la explicación de los conceptos básicos asociados a esta.

En tercer lugar (apartado 5), se profundizará en la relación entre mercado de valores y la inteligencia artificial mediante el análisis de estudios empíricos en los que se han utilizado modelos de inteligencia artificial como herramienta de inversión.

En última instancia (apartado 6), se recogerán las conclusiones del trabajo y se presentarán líneas de investigación futura recomendadas.

## 2. INTRODUCCIÓN A LA INVERSIÓN Y EL TRADING

El ámbito de la inversión se constituye como uno de los pilares fundamentales del universo financiero, repercutiendo de manera directa en la sociedad y en su desarrollo. A pesar de que en los últimos años han sido numerosas las incorporaciones y avances que han podido vislumbrarse en este ámbito, hay una serie de fundamentos básicos que se mantienen y que deben ser conocidos por todo aquel que busque adentrarse en esta área. Solamente de esta manera será posible asentar una base que permita la incorporación de herramientas más sofisticadas, como las que posteriormente se tratarán en este trabajo, y que permitirán ampliar el abanico de posibilidades del inversor.

En este sentido, uno de los conceptos más básicos es precisamente el **mercado de valores**. El mercado de valores corresponde a un “tipo de mercado de capitales en donde se opera sobre la renta variable y la renta fija, de forma estructurada, por medio de la compraventa de valores negociables” (Fernandez, 2019).

La combinación resultante de la compra y venta de determinados activos dentro del mercado de valores es lo que se conoce como **cartera**. El propósito general de todo inversor será construir una cartera que permita diversificar lo máximo posible el riesgo derivado de los activos y maximizar las posibles ganancias.

Como puede comprenderse, los inversores operan en el mercado de valores buscando obtener rentabilidad, y para ello llevan a cabo una serie de métodos determinados en función a sus preferencias y conocimientos. En términos generales, podemos hablar de que existen dos disciplinas ampliamente conocidas dentro de estos métodos: el análisis técnico y el análisis fundamental.

El **análisis fundamental** es un modelo de análisis de los mercados financieros basado en el estudio de tres aspectos fundamentales que afectan al precio de una acción: el entorno macroeconómico, la industria de la que forma parte la compañía y la propia situación de la compañía a varios niveles (financiero, administrativo...) (Shah, Isah, & Zulkernine, 2019).

Otra de las disciplinas más ampliamente utilizadas dentro del mundo de la inversión recibe el nombre de **análisis técnico**. El análisis técnico se centra en analizar tendencias

estadísticas derivadas de movimientos de precios, entre otros, para identificar oportunidades de inversión; las personas que llevan a cabo este tipo de estrategia tratan de identificar los patrones que se dan entre oferta y demanda, prestando atención a la información de precios y volumen disponible, para tratar de predecir los movimientos futuros de los activos (M. Harper, 2016). En definitiva, el análisis técnico se basa en el estudio del comportamiento de un activo y no en su valor. Normalmente, a los inversores que optan por operar realizando un análisis de los movimientos del mercado se les conoce como *traders* y acostumbran a operar en períodos de tiempo más cortos. De ahí la diferenciación en inversión y trading, aunque ambos términos se relacionan con la inversión *per se*.

Para llevar a cabo este tipo de disciplina, los inversores acostumbran a utilizar una serie de herramientas conocidas como indicadores técnicos que les permiten obtener más información sobre los activos que analizan.

Los **indicadores técnicos** son “cálculos heurísticos o matemáticos, basados en criterios como el precio o el volumen, que se utilizan en el análisis técnico” (Chen, Technical Indicator, 2019) con el fin de predecir el movimiento de un determinado mercado financiero y obtener beneficio (J. Murphy, 1999). Existen numerosas clases de indicadores técnicos según el criterio de análisis, pero escapa al objetivo de este trabajo presentar todos ellos. Sin embargo, sí que conviene conocer aquellos que se mencionarán con mayor frecuencia a lo largo de la investigación. Para ello, primero es conveniente conocer que existen cuatro tipos de indicadores técnicos: *momentum*, volatilidad, volumen y tendencia.

El término *momentum* en el entorno financiero hace referencia a la velocidad con la cual cambia el precio de un activo. En otras palabras, se refiere a la ratio de cambio de los movimientos del precio para un activo determinado (Hu, y otros, 2015).

Dentro de los indicadores técnicos que estudian el *momentum* encontramos el **oscilador estocástico** (*Stochastic Oscillator* o SO). En resumidas cuentas, lo que el oscilador estocástico realiza es una “comparación entre el precio de cierre de un activo y sus rangos de precio en un período de tiempo determinado” (Hayes, 2019). Este tipo de indicador acostumbra a representarse de la siguiente manera: %K.

Otro de los indicadores que analizan el *momentum* es el **índice de fuerza relativa** o RSI por sus siglas en inglés. Este índice mide la velocidad y la magnitud del cambio de precio de un activo (Chen, 2020).

En lo referente a los indicadores técnicos recogidos en la categoría *tendencia*, estos, como su propio nombre indica, sirven para estudiar la tendencia que ha tenido el precio de un determinado activo. Entre los más conocidos encontramos *Moving Averages* y *MACD*.

Los indicadores que estudian la *volatilidad* se centran en estudiar la magnitud de las variaciones de los precios. A mayor volatilidad, mayor será la magnitud del cambio del precio. Dentro de ellos encontramos **average true range** (ATR) y la desviación típica, entre otros (Cryptohopper, 2020)

Por último, los indicadores que estudian el *volumen* se centran en analizar el número de operaciones de compra y venta que se realizan para un determinado activo. El más conocido es *On Balance Volume* (Cryptohopper, 2020)

Tabla 1. Clases de indicadores técnicos y ejemplos

<i>Momentum</i>	<i>Tendencia</i>	<i>Volatilidad</i>	<i>Volumen</i>
<i>Stochastic Oscillator</i> <i>RSI</i>	<i>MACD</i> <i>Moving Averages</i>	<i>Average True Range</i> <i>Desviación Típica</i>	<i>On Balance Volume</i>

Como anticipábamos, el objetivo de todo inversor es obtener una rentabilidad determinada con sus operaciones en el mercado de valores. Para ello, una de las premisas básicas dentro del mundo del trading es la predicción del futuro movimiento de los precios. En este sentido, con el paso de los años y el desarrollo tecnológico se ha ido avanzando en esta materia, entrando en escena la utilización de la inteligencia artificial, tema principal de este trabajo y que se analizará y revisará en las páginas siguientes.

### 3. PREDICCIÓN DEL MERCADO DE VALORES

Dada la creciente relevancia de la inteligencia artificial de los últimos años, muchos han sido los académicos interesados en estudiar su aplicación en el contexto financiero. A pesar de que dentro del propio sector financiero encontramos diversas áreas en las que ha encontrado cabida la inteligencia artificial, nos resulta especialmente interesante hacer mención al ámbito de la inversión por las posibilidades que esta herramienta puede brindarnos.

Sin embargo, antes de comenzar a desarrollar las estrategias que apuestan por la utilización de inteligencia artificial en la inversión, debemos entender una serie de teorías que servirán de base para el planteamiento de nuestra investigación.

#### 3.1. Hipótesis del Mercado Eficiente

Dentro del mundo de los mercados financieros y la inversión existen varias teorías que apoyan o rechazan la posibilidad de batir al mercado de forma sostenida e intencionada. Una de las teorías más conocidas y aceptadas recibe el nombre de “Teoría del mercado eficiente”. Esta teoría se fundamenta en la idea de que no es posible llevar a cabo una estrategia capaz de batir al mercado debido a que los precios actuales de los activos ya reflejan la totalidad de la información disponible (Fama, 1991).

Entrando más en detalle, la teoría del Mercado Eficiente o *Efficient Market Hypothesis* sostiene que la propia naturaleza de los mercados modernos es eficiente, entendiendo este término como la situación en la que los precios ya incorporan toda la información relevante a dicho activo (Fama, 1991). Bajo esta situación, toda nueva información disponible afecta rápidamente al precio del activo y es incorporada en este sin demora. Si un activo recoge en su precio todos los posibles movimientos futuros, resulta imposible para un inversor obtener un beneficio superior al que obtendría mediante una selección aleatoria de los activos o mediante el uso del análisis técnico o fundamental (Malkiel, 2003).

A pesar de ser esta la teoría general, cabe destacar que existen puntualizaciones sobre la eficiencia de los mercados. Dentro de la propia teoría del mercado eficiente encontramos tres variantes: la forma débil, la cual considera que los precios se mueven de manera

aleatoria y que son independientes unos de otros, la forma semi-fuerte, que basa su planteamiento en que los precios se ajustan rápidamente de acuerdo con la información pública disponible, como puede ser el caso de dividendos, beneficios o eventos macroeconómicos, y, la última forma, conocida como fuerte, que afirma que los precios reflejan la totalidad de la información, ya sea pública o privada, por lo que nadie puede obtener una determinada ventaja en el mercado de valores (Shah, Isah, & Zulkernine, 2019).

### 3.2. Hipótesis del Paseo Aleatorio

Otra teoría ampliamente extendida y estrechamente relacionada con la anterior recibe el nombre de “Teoría del Paseo Aleatorio”. La teoría del paseo aleatorio, fuertemente popularizada en el siglo XX gracias al economista y académico Burton Gordon Malkiel, sostiene que: “los precios del mercado se mueven con el mismo indeterminismo práctico que las partículas que exhiben el movimiento browniano y que son, por lo tanto, equivalentes a nivel observacional a una serie aleatoria” (M. Harper, 2016, pág. 11). En otras palabras, los movimientos del mercado son totalmente aleatorios y no pueden ser previstos, por lo que no es posible desarrollar una estrategia gracias al análisis de los mismos; esta concepción iría en concordancia con la forma fuerte de la teoría del mercado eficiente. La concepción subyacente a esta teoría es que, al igual que veíamos en el caso de la teoría del mercado eficiente, la nueva información disponible afecta de forma directa a los mercados y queda reflejada inmediatamente en los precios de los activos; por este motivo, los precios de mañana solamente se verán afectados por las noticias de mañana y así sucesivamente. Dado el carácter intrínsecamente impredecible de las noticias, los movimientos de precios se convierten necesariamente en impredecibles y por tanto “aleatorios” (Malkiel, 2003).

Para dar explicación a los posibles beneficios puntuales obtenidos por determinados inversores, Burton Gordon Malkiel usaba la analogía de un conjunto de monos tirando dardos. Estos monos, que además deberían tener los ojos cerrados, lanzarían dardos a una lista de acciones. Cada mono lanzaría 20 dardos, seleccionando por tanto 20 activos. Pasado un año, los resultados mostrarían como la mitad de dichos monos habría conseguido obtener un rendimiento superior a la otra mitad (Malkiel, 1973) (M. Harper, 2016). Si se repitiese el mismo ejercicio al año siguiente volvería a darse el mismo resultado con los mismos monos u otros, sin embargo, en el caso de que algún mono



obtuviera rendimientos superiores a la media dos años seguidos, no dudaríamos en atribuir este suceso a la probabilidad y a la suerte más que a la elección inteligente de los activos (D. Arnott, Hsu, Kalesnik, & Tindall, 2013)

A pesar de que ambas teorías se apoyan en la idea de la imposibilidad de batir al mercado de forma estratégica, la diferencia entre las dos reside en que, la forma más débil de la teoría de mercado eficiente sostiene que sí existen ciertas asimetrías en el mercado que pueden ser utilizadas para obtener un beneficio superior en el corto plazo (M. Harper, 2016). Sin embargo, los precios futuros están determinados por información que no está presente en los precios anteriores. Por lo tanto, a pesar de que sea posible la obtención de ciertas ventajas puntuales que permitan la ganancia de un beneficio superior, la teoría del mercado eficiente sostiene que este tipo de ventajas será siempre puntual y que no es posible obtener de forma sistemática un beneficio superior mediante el análisis histórico de los precios (Fama, 1991).

En definitiva, ambas teorías están de acuerdo en afirmar que no es posible la elaboración de una estrategia capaz de obtener un beneficio superior al mercado de forma intencionada, puesto que ni el análisis histórico de los precios ni el análisis fundamental de las compañías serviría para predecir los movimientos futuros.

### 3.3. Introducción a la predicción del Mercado de Valores

A pesar de la amplia aceptación de las teorías del mercado eficiente y paseo aleatorio, a lo largo del siglo XX surge un gran interés a nivel internacional en tratar de predecir los movimientos del mercado. Comienzan por tanto a surgir estudios enfocados en encontrar explicaciones a los movimientos de los precios que puedan servir para predecir los movimientos futuros y así refutar las teorías del mercado eficiente y paseo aleatorio.

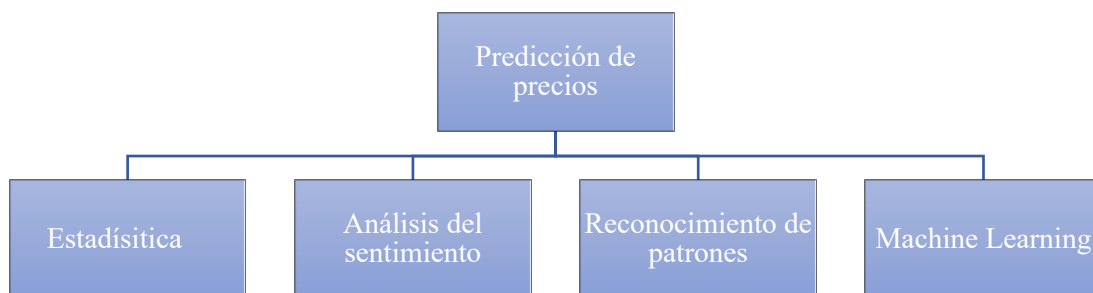
En esta línea, académicos y economistas a nivel global como Abu-Mostafa y Atiya, deciden centrar sus esfuerzos en estudiar si es posible desarrollar herramientas capaces de predecir los movimientos del mercado. Según los autores, la existencia de patrones en los precios junto con la presencia de correlaciones y eventos macroeconómicos no descontados hace surgir la posibilidad de presentar argumentos contra la teoría del mercado eficiente (Abu-Mostafa & Atiya, 1996). Por su parte, otros autores también hacen hincapié en el análisis de elementos psicológicos y conductuales, pues consideran

que muchas de las reacciones humanas ante determinadas situaciones son repetitivas y que, por lo tanto, podrían llegar a preverse (Malkiel, *The Efficient Market Hypothesis and Its Critics*, 2003).

En definitiva, se trata de realizar un análisis multidisciplinar capaz de encontrar patrones dentro de los mercados que expliquen su desempeño futuro. Este terreno ha evolucionado enormemente desde sus inicios, incluyendo cada vez más técnicas y elementos, pero, en términos generales, basa su actividad en la predicción de los movimientos del mercado mediante la utilización del análisis técnico.

De acuerdo con Shah, Isah y Zulkernine (2019) dentro de la propia predicción de precios encontramos cuatro categorías: predicciones estadísticas, reconocimiento de patrones, machine learning y análisis del sentimiento. Estas categorías, por lo general, pueden agruparse dentro del análisis técnico previamente mencionado.

Ilustración 1. Categorías incluidas en la predicción de precios



Fuente: elaboración propia con datos de Shah, Isah y Zulkernine (2019)

En un primer momento, los modelos más utilizados para tratar de predecir los movimientos de los precios en el mercado eran los estadísticos, y más concretamente, de regresión (Gharoie Ahangar, Yahyazadehfar, & Pournaghshband, 2010). Sin embargo, el desarrollo de la tecnología y la consecuente mayor capacidad de procesamiento de la información ha abierto una nueva posibilidad a la hora de predecir los movimientos de los activos. En la actualidad, han surgido formas más avanzadas de predicción basadas en la utilización de inteligencia artificial que han hecho aumentar la popularidad de esta tendencia.

Antes de profundizar en cómo las herramientas de inteligencia artificial son utilizadas en la predicción del mercado de valores, debemos hacer un breve repaso de los modelos utilizados anteriormente con el fin de poder evaluar si se ha producido una mejora de la predicción como consecuencia del desarrollo de este tipo de tecnología.

### 3.4. Modelos estadísticos para la predicción del mercado de valores

En un primer momento, los estudios enfocados en la predicción del mercado de valores se centraron en el uso de herramientas estadísticas, mayormente basadas en modelos de regresión lineal. Los modelos lineales resultan fáciles de desarrollar e implementar, además de ser relativamente sencillos de interpretar, por lo que resultaron realmente interesantes a la hora de avanzar en esta materia (Gharoie Ahangar, Yahyazadehfar, & Pournaghshband, 2010)

Dentro de los modelos estadísticos encontramos diferentes categorías tales como Auto-Regressive Moving Average (ARMA), Auto-Regressive Integrated Moving Average (ARIMA), Generalized Autoregressive Conditional Heteroskedastic (GARCH) volatilidad o el modelo Smooth Transition Autoregressive (STAR) (Shah, Isah, & Zulkernine, 2019). Entrar a evaluar cada uno de los modelos mencionados está fuera de los propósitos de este trabajo, pero nos es posible observar que las principales técnicas dentro de los modelos de predicción estadísticos están basadas en el análisis de regresión.

Sin embargo, uno de los problemas que los modelos de regresión lineal plantean a la hora de predecir de manera precisa los movimientos del mercado de valores reside en su poca flexibilidad a la hora de adaptarse de manera autónoma. Debido al gran número de factores que influyen en las series temporales de los mercados, estas incluyen una gran cantidad de ruido, volviéndose prácticamente aleatorias y adquiriendo un carácter no lineal (Gharoie Ahangar, Yahyazadehfar, & Pournaghshband, 2010).

Estos modelos de regresión, al estar basados en determinados modelos lineales, no permiten un ajuste automático de los parámetros, por lo que las predicciones no terminan de ser del todo correctas cuando se aplican en circunstancias de no linealidad (Pakdaman Naeini, Taremian, & Baradaran Hashemi, 2010). Las limitaciones que este tipo de modelos encuentran a la hora de trabajar con relaciones no lineales hace muy complicado su uso para la predicción de los movimientos del mercado de valores, que son cada vez

más complejos y tienden a caracterizarse por la no linealidad y el dinamismo (Gharoie Ahangar, Yahyazadehfar, & Pournaghshband, 2010).

Por tanto, las investigaciones centradas en la predicción del mercado de valores mediante el uso de modelos estadísticos en general, y de regresión lineal en particular, han tomado diversas formas, pero muchas de ellas han terminado por reforzar la hipótesis del mercado eficiente. El análisis de la literatura nos indica que la propia naturaleza del mercado de valores, caracterizada por la complejidad y la no linealidad, no permite que estas técnicas funcionen correctamente en la práctica, por lo que otras herramientas podrían resultar más interesantes a la hora de tratar de predecir los movimientos del mercado.

#### **4. INTELIGENCIA ARTIFICIAL EN LA PREDICCIÓN DEL MERCADO DE VALORES**

Una vez conocidos los fundamentos básicos tanto de la inversión como de la predicción del mercado de valores, para poder continuar con el desarrollo de este trabajo, resulta imprescindible introducir los conceptos ligados a la inteligencia artificial que se aplicarán en las páginas siguientes. De esta manera, se podrá comprender posteriormente cómo la inteligencia artificial se relaciona con el mercado de valores, constituyéndose como una herramienta de gran utilidad en el ámbito de la predicción de precios.

##### **4.1. Introducción a la Inteligencia Artificial**

El término **Inteligencia Artificial** (IA) es posiblemente uno de los más utilizados hoy en día, sin embargo, cuenta con más de 60 años de existencia, pues ya en 1956 John McCarthy lo definía como *“la ciencia e ingeniería de hacer máquinas inteligentes”* (Peart, 2017). Más recientemente, concretamente en el año 2017, el Consejo de Estabilidad Financiera acuñaría el concepto como *“la teoría y el desarrollo de sistemas computacionales capaces de llevar a cabo tareas que tradicionalmente requerían de inteligencia humana”* (Bonnie G. Buchanan, 2019).

A pesar de la diversidad de definiciones, todas están de acuerdo en afirmar que la inteligencia artificial se centra en determinadas capacidades autónomas de los sistemas computacionales, por lo que, en términos generales, podemos decir que la inteligencia artificial engloba una gran variedad de tecnologías informáticas cuyo objetivo es la

“creación de sistemas capaces de funcionar de forma independiente” (Gautam, Gupta, Awasthi, & Gautam, 2019).

Debido a la amplitud del término, dentro de la inteligencia artificial encontraríamos varios enfoques o categorías en las que esta se aplica con un fin determinado. Dentro de estos enfoques encontraríamos *machine learning*.

**Machine Learning** (ML) se define como un enfoque dentro de la IA en el cual, ante nuevos escenarios, es posible la toma de decisiones por parte de la máquina mediante la extracción de datos pasados y la configuración de algoritmos capaces de aprender de su propia experiencia (Bonnie G. Buchanan, 2019) En otras palabras, *machine learning* es una disciplina que, mediante algoritmos, elabora modelos capaces de analizar datos y aprender de ellos para elaborar conclusiones.

A pesar de que existen ciertas categorías intermedias, podemos hablar de dos clases generales dentro de *machine learning*: el aprendizaje supervisado o *supervised learning*, sobre el cual se estructurará este trabajo, y el aprendizaje no supervisado (*unsupervised learning*). La diferencia entre ambos reside en que, en el aprendizaje supervisado, la información que se introduce para el entrenamiento del algoritmo está categorizada, circunstancia que no se da en el segundo caso (Shah, Isah, & Zulkernine, 2019).

### *Aprendizaje supervisado*

Con el fin de facilitar la comprensión de este trabajo al lector, procederemos a explicar de forma simplificada cómo se estructura el proceso de aprendizaje supervisado en *machine learning*.

En primer lugar, debe seleccionarse una base de datos sobre la cual se estructurará el modelo que queremos elaborar. Dentro de la base de datos seleccionada deben crearse dos grupos: el conjunto de entrenamiento y el conjunto de prueba. El conjunto de entrenamiento, también conocido como *training set*, recogerá las variables que serán utilizadas para el entrenamiento del algoritmo. Por su parte, el conjunto de prueba o *test set*, servirá para probar en qué medida los resultados obtenidos por el modelo se corresponden con los datos reales (Google Developers, 2020).

Realizada la división de la base de datos en dos grupos, las variables de entrenamiento (*input variables*) deben ser introducidas en el modelo. Una vez introducidas, el algoritmo analizará los patrones que en estas se dan, así como las correlaciones entre ellas, para obtener un resultado (*output*). A medida que el algoritmo vaya analizando mayor cantidad de información será capaz de mejorar la precisión del resultado obtenido, pues una de las principales ventajas de *machine learning* es la posibilidad que tienen las máquinas de aprender gracias a su propia experiencia y entrenamiento (Gautam, Gupta, Awasthi, & Gautam, 2019).

Ilustración 2. Simplificación de un modelo



Fuente: elaboración propia

El aprendizaje supervisado puede a su vez clasificarse en problemas de *regresión* y de *clasificación* (Google Developers, 2020):

*Problemas de clasificación:* En los problemas de clasificación el resultado es una categoría entre un número limitado de categorías (Martinez Heras, 2020). Por ejemplo, si aplicásemos este tipo de problemas a la predicción de la dirección del IBEX 35 mañana la pregunta a realizar sería: ¿bajará el IBEX 35 mañana? Y la respuesta quedaría recogida en una de las dos opciones: “sí” o “no”. Por otro lado, un modelo de clasificación también predice la probabilidad de que se de cada una de las opciones; si, por ejemplo, queremos conocer si el mensaje que recibimos es spam y el modelo nos devuelve un 0.99, podemos considerar que muy probablemente el mensaje sea spam. La consideración dentro de una categoría u otra (spam o no spam, por ejemplo) dependerá del umbral de clasificación que definamos; para el ejemplo anterior, un valor por encima de ese umbral sería considerado spam (Google Developers, 2020).

*Problemas de regresión:* En el caso de los problemas de regresión, el resultado obtenido es un número concreto (Martinez Heras, 2020). Utilizando el ejemplo anterior, la pregunta a realizar sería ¿cuánto bajará mañana el IBEX 35? Siendo la respuesta un número concreto.

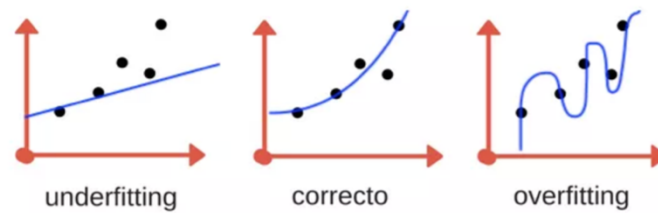
### *Conceptos asociados a los modelos de Machine Learning*

**Generalización.** El concepto *generalización* dentro del campo de *machine learning* hace referencia a cómo los conceptos aprendidos por un modelo basado en *machine learning* se aplican a ejemplos específicos que no han sido vistos por el modelo en su entrenamiento (Google Developers, 2020). El objetivo de un buen modelo de *machine learning* será generalizar bien nuevos datos basándose en los datos de los que ha aprendido, pues esto permitirá al modelo hacer predicciones con información que no ha visto nunca. Muy relacionado con este término encontramos el concepto de *overfitting* o sobreajuste y *underfitting* o subajuste (Grosse, 2017).

**Overfitting/ Sobreajuste.** De acuerdo con Grosse (2017), “consideramos que un modelo tiene un problema de sobreajuste cuando este es capaz de funcionar correctamente con los datos de entrenamiento, pero falla a la hora de generalizar”. En otras palabras, el espacio que hay entre el conjunto de entrenamiento y el conjunto de prueba es demasiado grande, por lo que las predicciones no son precisas (Grosse, 2017).

**Underfitting/ Subajuste.** El problema de subajuste implica que el modelo no funciona correctamente ni con el conjunto de entrenamiento ni con el conjunto de prueba. Este caso acostumbra a venir derivado de la insuficiencia de datos de entrenamiento, circunstancia que no permite al modelo extraer conclusiones y correlaciones (Google Developers, 2020).

Ilustración 3. Representación gráfica de la capacidad de generalización de un modelo de Machine Learning



Fuente: (Bagnato, 2017)

**Selección de variables relevantes:** Uno de los elementos claves para la formulación de un modelo cuyo objetivo es la predicción es la correcta selección de las variables de entrenamiento de dicho modelo. La selección de variables, como su propio nombre indica, consiste en la elección de aquellas variables que servirán como *input* para el modelo que se busca elaborar de entre los datos seleccionados. Por ejemplo, si los datos que se han obtenido corresponden al IBEX 35, podrían seleccionarse como variables de entrenamiento el precio de cierre y el precio medio u optar por utilizar indicadores técnicos elaborados con estos datos.

**Medidas de precisión:** De igual manera que resulta imprescindible seleccionar correctamente las variables de entrenamiento y llevar a cabo el entrenamiento del modelo para poder obtener un output, es necesario poder evaluar la precisión que este ha tenido a la hora de predecir los resultados. Para ello, hace falta distinguir el tipo de problema ante el que nos encontramos (página 22).

Problemas de regresión: Entre las medidas de precisión más utilizadas en los problemas de regresión encontramos *mean absolute error* y *mean squared error*.

- *Mean absolute error* es la “media de la magnitud de errores que se dan en un set de predicciones” (Teshfamariam & Liu, 2013).
- *Mean squared error* mide la media de los cuadrados de los errores mediante la medición de la distancia entre los puntos y la línea de regresión. Esta medida de error elimina los signos negativos y da un peso mayor aquellas muestras en las que el modelo se ha equivocado más (Universidad de Pensilvania, 2020).



Problemas de clasificación: Por otro lado, dentro de los problemas de clasificación las medidas de precisión más utilizadas son las curvas ROC, así como los parámetros exactitud (*accuracy*), precisión (*precision*), sensibilidad (*sensitivity*), especificidad (*specificity*) y F-Score.

- Exactitud (*accuracy*): es la fracción entre el número correcto de predicciones y el total de predicciones (Google Developers, 2020).
- Precisión (*precision*): es la división entre el número de verdaderos positivos (*true positives*) y la suma de verdaderos positivos y falsos negativos (*false negatives*) (Google Developers, 2020).
- Sensibilidad (*sensitivity o recall*): es la ratio de positivos que han sido correctamente identificados como tal (Google Developers, 2020).
- Especificidad (*specificity*): es la ratio de negativos que han sido correctamente identificados como tal (Cai, 2014)
- F-Score: es la media armónica ponderada de la precisión y la sensibilidad (Ghoneim, 2019).
- Las curvas ROC se representan gráficamente y sirven para indicar el desempeño de un modelo de clasificación ponderando la sensibilidad y especificidad del modelo para distintas probabilidades de corte (el umbral de clasificación previamente mencionado) (Google Developers, 2020).

Para más información acerca de las formulas que componen los parámetros anteriores, ver Anexo I.

#### 4.2. Aplicación de la Inteligencia Artificial al Mercado de Valores

El avance tecnológico y la consecuente aparición de la inteligencia artificial a finales del siglo XX han permitido continuar investigando en el ámbito de la predicción de precios. Este tipo de tecnología, que también está siendo utilizada en otras áreas tales como la aviación o la sanidad, ha encontrado un gran mercado para desarrollarse en el mundo de la inversión. Si anteriormente se utilizaban modelos estadísticos para tratar de prever los futuros movimientos del mercado, el surgimiento de estas nuevas herramientas de inteligencia artificial, como es el caso de *machine learning*, ha permitido el desarrollo de modelos más complejos que se presentan como superiores a los utilizados hasta la fecha.

Como previamente se adelantaba, la inteligencia artificial, y más concretamente el *machine learning*, permite a las máquinas procesar enormes cantidades de información y extraer conclusiones que serían mucho más difíciles, o incluso imposibles, de obtener para un ser humano. El funcionamiento de estas herramientas está basado en el desarrollo de algoritmos capaces de aprender de la información que se les aporte (*inputs*) para ofrecer un resultado (*output*) (Gautam, Gupta, Awasthi, & Gautam, 2019).

Las características propias de la inteligencia artificial, que hacen posible la configuración de máquinas capaces de analizar información y encontrar patrones, han abierto un nuevo frente contra las hipótesis del mercado eficiente y paseo aleatorio. Desde su surgimiento, académicos de todo el mundo han centrado sus esfuerzos en la elaboración de modelos que, mediante la configuración de algoritmos, son capaces de aprender de la información disponible en el mercado y pueden extraer conclusiones en función de lo que se espera de los precios. Como previamente se anticipaba, estos modelos se construyen en base a algoritmos, por lo que la elección del tipo de algoritmo a utilizar adquiere una gran importancia. Cada clase de algoritmo tiene su propia forma de aprender patrones y predecir los resultados futuros, por lo que la elección de uno u otro será muy influyente en el resultado final del modelo.

Entre los algoritmos utilizados para la predicción de precios cabe destacar, por su frecuente utilización, *Support Vector Machine (SVM)*, *Artificial Neural Networks (ANN)* y *Random Forest*. Sin embargo, dentro de estas categorías encontramos todavía más variantes, lo que nos hace tener que elegir entre numerosas posibilidades.

La diversidad de algoritmos que pueden ser utilizados con este fin ha abierto una interesante discusión sobre cuál es el algoritmo más preciso a la hora de predecir los movimientos de los precios. Durante los últimos años, investigadores y economistas de todo el mundo han redactado estudios utilizando diferentes algoritmos para analizar su precisión. Sin embargo, el gran número de estudios elaborados y la diversidad de modelos utilizados hace surgir la necesidad de recopilar la información disponible para que esta pueda ser presentada de forma clara y comparable.

Por tanto, con el fin de conocer más ampliamente este tipo de herramientas y su efectividad, se explicarán a continuación algunas de las clases de algoritmos más

utilizadas para la predicción de los mercados de valores. Por otro lado, también se prestará especial atención a los inputs utilizados y a los resultados obtenidos. De esta manera, será posible reunir buena parte la información disponible sobre los distintos tipos de algoritmos utilizados en la predicción del mercado de valores, de forma que sea posible una comprensión de los mismos sin importar la rama de conocimiento del lector.

#### 4.2.1. Artificial Neural Networks

La utilización de redes neuronales para la predicción comienza con el trabajo de Hu en 1964, quien decidió emplear esta herramienta para hacer predicciones climatológicas. Hoy en día, las redes neuronales han ampliado su rango de uso a muchas más áreas, entre ellas, la inversión (Pakdaman Naeini, Taremian, & Baradaran Hashemi, 2010).

Las redes neuronales artificiales o *Artificial Neural Networks* son modelos computacionales inspirados en el funcionamiento del cerebro humano que tratan de simular nuestra red neuronal y las conexiones que en ella se producen para extraer un resultado (M. Harper, 2016). Estas redes neuronales están formadas por neuronas conectadas entre sí mediante operaciones matemáticas, de manera que se constituyen en bloques. Los bloques de redes neuronales reciben el nombre de “*layers*” o capas y al conectarse entre sí dan lugar a redes cada vez más complejas. Cuantas más capas tenga una red neuronal más compleja será y más información podrá analizar antes de elaborar una respuesta (Thmaz & C.Medin, 2019). Existen tres tipos de capas: de entrada, de salida y ocultas. La capa de entrada recibe la información del entorno, las capas ocultas son aquellas que no tiene conexión directa con el entorno, por último, la capa de salida está compuesta por aquellas neuronas que proporcionan la respuesta de la red neuronal. (Larrañaga, Inza, & Moujahid, 1997).

En los últimos años, las redes neuronales artificiales han ganado gran popularidad como herramientas útiles en problemas de predicción debido a las posibilidades que brindan. Por un lado, la literatura disponible nos permite conocer que las redes neuronales son capaces de aproximar prácticamente cualquier función no lineal con suficiente precisión, por lo que son mucho mejores que los modelos econométricos lineales tradicionales a la hora de descubrir relaciones complejas entre componentes de muchas series temporales (M. Harper, 2016) (Dharmaraja, Vineet, & Abhishek, 2019). En otras palabras, las redes neuronales tienen una gran capacidad para reconocer patrones y clasificar datos incluso

en situaciones complicadas. Si tenemos en cuenta las cualidades que definen el mercado de valores, fuertemente caracterizado por la gran cantidad de información que en este se genera y la complejidad de la misma, encontraremos que la aplicación de redes neuronales en este ámbito puede ser muy acertada. De hecho, la capacidad de estos algoritmos de aprender de datos no lineales, ya han permitido su amplia utilización para la resolución de problemas de predicción en muchas otras áreas (Kimoto, Asakawa, & Yoda, 1990)

Por otro lado, las redes neuronales no contienen fórmulas estándar, lo que les da la habilidad de auto-ajustarse, siendo capaces de resolver problemas con muy poco conocimiento sobre el modelo. Además, es posible para este tipo de sistemas reconocer nuevos patrones incluso sin estar estos previamente presentes en las bases de datos utilizadas para su entrenamiento (Pakdaman Naeini, Taremian, & Baradaran Hashemi, 2010), lo que permitiría su aplicación en la predicción de precios.

El funcionamiento de este tipo de sistemas, tal y como veíamos anteriormente, se basa en la introducción de una serie de variables (inputs) que servirán para entrenar al modelo y que este elabore una respuesta (output). En este sentido, es muy importante tener en cuenta que, tanto la elección de variables de entrenamiento, como el método de entrenamiento en sí, será decisivo en el resultado final del modelo.

Dentro de las redes neuronales también encontramos diferentes tipos; es importante tener esto presente puesto que no todos son capaces de predecir series temporales. Una de las más utilizadas con esta finalidad recibe el nombre de Multilayered Feed-Foward Neural Network (MFFNN). Sin entrar en detalle, pues escapa al propósito de este trabajo, debemos conocer que los estudios nos indican que las redes neuronales del tipo MFFNN han demostrado ser bastante eficientes a la hora de hacer aproximaciones de funciones (Samek & Varacha, 2013) por lo que pueden ser una buena elección a la hora de tratar de predecir el mercado de valores. Otro tipo de red neuronal interesante para este objetivo son las Redes Neuronales Recurrentes. Este tipo de redes, a pesar de ser más complejas que las anteriores, son potencialmente más precisas debido a su mayor similitud con las neuronas reales (M. Harper, 2016); dentro de esta clase encontraríamos las redes Elman y las redes Jordan, a las cuales haremos mención posteriormente.

Sin embargo, las redes neuronales también tienen sus desventajas. Debido a la caracterización de estas de modelos estadísticos no paramétricos, no se prestan a la interpretación (M. Harper, 2016). Esto quiere decir que, a pesar de que son altamente capaces de descubrir relaciones no lineales, no proporcionan casi información sobre la naturaleza de dichas relaciones, lo que se traduce en una dificultad para explicar el valor de los modelos. Esta circunstancia puede suponer un problema a la hora de extrapolar las predicciones de un modelo a otro. Por tanto, la capacidad de predicción de las redes neuronales dependerá de la similitud entre el proceso de generación de datos desconocidos en vigor durante el momento en el que se entrenó la red neuronal y los intervalos futuros para los que se desea hacer la predicción (M. Harper, 2016) A menor similitud de los datos, menor fiabilidad tendrá la predicción realizada por la red neuronal.

Por otro lado, el problema de *over-fitting* o sobreajuste es otra de las limitaciones que se encuentran presentes en los modelos creados mediante ANN. Debido al gran número de parámetros que deben fijarse para el correcto funcionamiento de las redes neuronales, es más fácil caer en problemas de sobreajuste, pues encontrar el número óptimo de unidades para que el modelo sea capaz de generalizar correctamente se vuelve todavía más complicado (Hegazy, Soliman, & Abdul Salam, 2013) (Wanjawa & Muchemi, 2014).

De igual manera, es importante tener en consideración una serie de factores esenciales a la hora de utilizar las redes neuronales para este tipo de problemas de predicción; de entre ellos cabe destacar la estructura de las redes, los datos que se van a utilizar para entrenar el algoritmo o los procedimientos de evaluación que se van a utilizar para medir el desempeño de este, pues serán muy influyentes en la precisión del modelo.

#### 4.2.1. Support Vector Machines

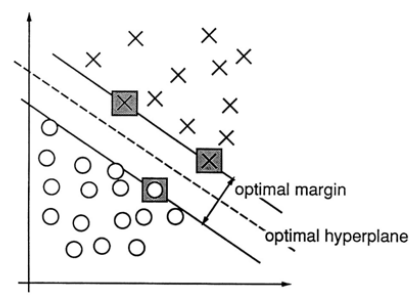
El siguiente tipo de algoritmo que se explicará en este trabajo recibe el nombre de *Support Vector Machines*, también conocido como SVM. *Support Vector Machines* es conocido por ser uno de los modelos más utilizados con la finalidad de predecir los movimientos del mercado de valores, siendo considerado por muchos el mejor modelo para resolver este tipo de problemas.

*Support Vector Machines* tiene la característica de ser un clasificador lineal binario no probabilístico, por lo que este tipo de algoritmo puede ser utilizado en problemas de

clasificación y regresión. Como ocurría en el caso anterior, la obtención de resultados mediante SVM se fundamenta en la elaboración de un modelo, en el cual se introducen una serie de datos de entrenamiento de los que este es capaz de aprender y emitir un resultado (Cortes & Vapnik, 1995).

En el caso de los problemas de clasificación, que son aquellos en los que nos centraremos, lo que se trata de encontrar es el hiper plano óptimo que separa dos clases. Entrando más en detalle, el funcionamiento de esta clase de algoritmos se basa en trazar cada componente de los datos como un punto en el espacio n-dimensional (siendo n el número de elementos de la base de datos disponible) siendo el valor de cada elemento el valor de una coordenada particular y, por lo tanto, la clasificación se realiza encontrando el hiper plano que diferencia las dos clases de forma explícita (Cortes & Vapnik, 1995).

Ilustración 4. Representación SVM



Fuente: (Cortes & Vapnik, 1995)

Revisando la literatura disponible encontramos algunas de las ventajas que harían interesante la utilización de este tipo de algoritmos para los problemas de predicción. En primer lugar, los algoritmos del tipo SVM están basados en el principio de minimización del riesgo estructural, que incorpora la capacidad de prevenir el problema de *over-fitting* previamente mencionado. Por este mismo motivo, este tipo de algoritmos han demostrado tener una alta capacidad de generalización (Sampat Patil, Patidar, & Jain, 2016).

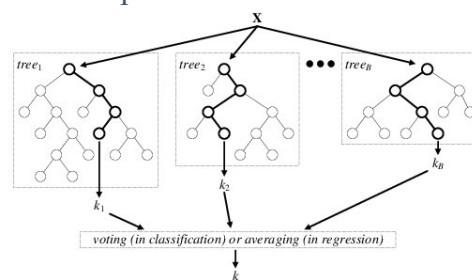
Por otro lado, como anteriormente se mencionaba, la gran cantidad de ruido que hay en el mercado puede ser un impedimento para el funcionamiento de los modelos predictivos, por lo que es importante la reducción de este y la correcta preparación de los datos antes de llevar a cabo cualquier predicción. Anteriormente se mencionaban las limitaciones a las que se enfrentan los modelos basados en ANN, los algoritmos SVM por el contrario,

consiguen superar estas limitaciones gracias a los sólidos fundamentos teóricos de la teoría VC: mientras que las soluciones aportadas por los modelos ANN son locales, las soluciones generadas a través de SVM son óptimos globales (Hegazy, Soliman, & Abdul Salam, 2013). Además, los algoritmos SVM son capaces de funcionar bien con bases de datos grandes (Sampat Patil, Patidar, & Jain, 2016), cualidad importante para poder trabajar correctamente con series temporales de precios.

#### 4.2.1. Random Forest

El último tipo de algoritmo que analizaremos recibe el nombre de *Random Forest*. *Random Forest* es un tipo de algoritmo compuesto por múltiples árboles de decisión cuyo *output* recoge el conjunto de salidas de los árboles de decisión individuales (Khaidem, Saha, & Roy Dey, 2016).

Ilustración 5. Simplificación modelo Random Forest



Fuente: (Verikas, Vaiciukynas, Gelzinis, Parker, & Olsson, 2016)

En este caso, los árboles de decisión operan de forma conjunta, pero cada uno es entrenado en un espacio diferente, de forma que no están correlacionados. Esto también implica que ninguno de los árboles recibe la totalidad de los datos de entrenamiento. De esta manera, a pesar de que el *bias* del modelo se incrementa, el ruido que puede haber en las variables introducidas (*inputs*) no causa alteraciones significativas (Khaidem, Saha, & Roy Dey, 2016). En definitiva, la idea detrás de este modelo se basa en la creencia de que una serie de modelos no correlacionados operando juntos actúa mejor que cualquiera de ellos por separado, pues los árboles de decisión se protegen entre ellos de sus errores individuales.

Otra de las ventajas que podemos encontrar a la hora de utilizar el modelo *Random Forest* guarda relación con el problema de sobreajuste (*over-fitting*). Los estudios realizados nos indican que la ratio de error OOB, utilizada para estimar el error de generalización, tiende

a converger a medida que el número de árboles de decisión aumenta. Esto quiere decir que, aunque aumentemos el número de árboles de decisión, *Random Forest* no reporta problemas de sobreajuste (Basak, Saha, Kar, Khaidem, & Dey, 2019). Sin embargo, una de las desventajas de la utilización de este algoritmo reside en la dificultad que puede haber a la hora de entender los resultados producidos por este (Khaidem, Saha, & Roy Dey, 2016). Por ello, si lo que buscamos es entender las relaciones que hay entre los datos y los resultados, puede no ser el algoritmo más indicado.



## 5. ANÁLISIS DE ESTUDIOS

En este apartado, se analizarán estudios empíricos en los que se hayan utilizado modelos de *machine learning* para predecir el mercado de valores. Para comenzar, se analizarán los modelos *Artificial Neural Networks*, posteriormente, se estudiarán los modelos del tipo *Support Vector Machines* y, por último, se investigarán los modelos *Random Forest*. Se prestará atención a los datos utilizados, las variables de entrenamiento seleccionadas, el tipo de modelo elaborado y los resultados obtenidos.

Tabla 2. Resumen estudios basado en la predicción mercado de valores con IA

Artículo	Modelo Utilizado	Inputs	Output	Medidas de precisión	Clasificación del problema
<b>Est. 1 (Pakdaman Naeini, Taremian, &amp; Baradaran Hashemi, 2010)</b>	MLP y Elman	Precio mínimo, máximo y medio	Mercado de valores de Teherán	MAE, MSE, RMSE, MAD	Regresión y clasificación
<b>Est. 2 (M. Harper, 2016)</b>	MLP	Precio de cierre y valor de la prima	S&P 500, Dow Jones, NASAQ 100	RMSE, Coeficiente Correlación de Pearson	Regresión
<b>Est. 3 (Qiu &amp; Song, 2016)</b>	ANN + GA	Indicadores técnicos (Tablas 7 y 8)	Nikkei 225 Index	Hit ratio	Clasificación
<b>Est. 4 (Madge, 2015)</b>	SVM	<i>Momentum y volatilidad</i>	NASDAQ 100	<i>mean prediction accuracy</i>	Clasificación
<b>Est. 5 (Di, 2014)</b>	SVM + Extremely Randomized Tree	Indicadores técnicos Tabla 11	Apple, Microsoft, Amazon	<i>Accuracy</i>	Clasificación
<b>Est 6. (Yu, Wang, &amp; Keung Lai, 2005)</b>	SVM + GA	Indicadores técnicos Tabla 13	S&P 500	Hit ratio	Clasificación
<b>Est. 7 (Khaidem, Saha, &amp; Roy Dey, 2016)</b>	RF + <i>exponential smoothing</i>	Indicadores técnicos Tabla 17	Apple, General Electric Samsung	Curvas ROC	Clasificación
<b>Est. 8 (Bafandeh Inmandoust &amp; Bolandraftar, 2014)</b>	RF	Indicadores técnicos y fundamentales Tabla 18	Mercado de valores de Teherán	<i>Accuracy</i>	Clasificación
<b>Est. 9 (Basak, Saha, Kar, Khaidem, &amp; Dey, 2019)</b>	RF + <i>exponential smoothing</i>	Indicadores técnicos: RSI, SO, WPR, MACD, PROC, OBV	Apple, Toyota (+8)	F-score, precisión, exactitud y especificidad	Clasificación

## 5.1. Artificial Neural Networks

### **Estudio 1 – Pakdaman Naeini, Taremian y Baradaran Hashemi (2010)**

#### *Modelo de predicción utilizando MLP, Elman RN y Regresión Lineal*

El primer estudio que analizaremos corresponde a los autores Pakdaman Naeini, Taremian y Baradaran Hashemi. En este caso, el trabajo de investigación se basa en una comparativa a dos niveles: por un lado, entre diferentes modelos de redes neuronales y, por otro lado, de estas redes neuronales con un modelo de regresión lineal. El fin último de este estudio se centra en determinar cuál de los modelos es más preciso a la hora de predecir el valor y la dirección del mercado. Por tanto, el problema es considerado tanto de regresión como de clasificación

Los datos utilizados para este estudio fueron obtenidos de 1094 compañías incluidas en el mercado de valores de Teherán entre el año 2000 y el año 2005. Concretamente, fueron utilizados como variables o *inputs* los precios, mínimos, máximos y medios de estas compañías durante el período de tiempo mencionado.

En lo referente al modelo utilizado, se utilizaron dos redes neuronales, Multilayer Perceptron feed-forward y Elman recurrent; para el entrenamiento de dichas redes se utilizó el algoritmo Back Propagation, concretamente, Levenberg-Marquardt. Por su parte, también se elaboró un modelo de regresión lineal que serviría para comparar los resultados obtenidos y discutir el posible mejor desempeño de las redes neuronales a la hora de predecir los movimientos del mercado. Con la elaboración de estos modelos se pretendía, en primer lugar, predecir la dirección del mercado y posteriormente, el valor futuro de la acción analizando su comportamiento pasado.

En este caso, se utilizaron varias herramientas para evaluar la precisión del modelo. En primer lugar, se analizó la desviación media (*mean absolute deviation*), el error porcentual absoluto medio (*mean absolute percentage error*), el error cuadrático medio (*mean squared error*) y la raíz del error cuadrático medio (*root mean squared error*). Junto con estas medidas, también se estudió el porcentaje de predicciones correctas mediante (1). Por otro lado, debido a que en este caso también se buscaba predecir la cuantía de la variación del precio y no únicamente el movimiento del mercado, se utilizaron dos medidas adicionales (2) (3)

$$(1) \text{ Correct Forecast Trend} = \frac{1}{|\text{Validation Set}|}$$

$$(2) \frac{\text{Correct Forecast Trend}}{\text{Real Trend}}$$

$$(3) \frac{\text{Incorrect Forecast Trend}}{\text{Real Trend}}$$

Los resultados obtenidos muestran que, de entre los dos métodos con redes neuronales utilizados, las predicciones realizadas por Elman Recurrent Network fueron más precisas que las realizadas con MLP. Sin embargo, el estudio también nos indica que el modelo de regresión lineal es igualmente superior a la hora de predecir la dirección de los precios que MLP. Por tanto, en lo referente a la predicción de la dirección, ninguno de los métodos que emplearon redes neuronales (Elman y MLP) resultaron ser mejores que el modelo de regresión lineal.

Tabla 3. Modelo ANN 1

<b>Artificial Neural Networks</b> - (Pakdaman Naeini, Taremian, & Baradaran Hashemi, 2010)	
Clasificación del problema	Problema de regresión y clasificación
Input	Precio mínimo, precio máximo y precio medio
Modelo empleado	MLP y Elman Recurrent Network / Modelo de regresión lineal
Output	Mercado de valores de Teherán
Medidas de precisión	<i>Mean Absolute Deviation, Mean Absolute Percentage Error, Mean Squared Error y Root Mean squared Error</i>
Resultados	<u>Predicción de la dirección:</u> Regresión Lineal y Elman RN > MLP <u>Predicción de la cuantía de la variación:</u> MLP > Regresión Lineal y Elman

Fuente: elaboración propia con datos de (Pakdaman Naeini, Taremian, & Baradaran Hashemi, 2010)

Sin embargo, los autores también pudieron observar que, a la hora de predecir la cuantía de los cambios de precios, el error observado es menor cuando se utiliza MLP. De hecho, pudo observarse que cuando MLP acierta en su predicción del movimiento del mercado, la cantidad de variación que predice es muy cercana a los datos reales.

En conclusión, los resultados del estudio llevado a cabo por Pakdaman, Taremian y Baradaran nos indican que los modelos elaborados mediante redes neuronales no pueden considerarse mejores que el modelo de regresión lineal a la hora de predecir la dirección de los movimientos del mercado. Sin embargo, cuando lo que se analiza es la precisión a la hora de predecir la cuantía de las variaciones, el grado de error del modelo MLP es menor al de los otros dos métodos analizados (Pakdaman Naeini, Taremian, & Baradaran Hashemi, 2010)

## **Estudio 2 – M. Harper (2016)**

### Modelo de predicción utilizando MLP

El siguiente estudio por analizar fue realizado por M. Harper en el año 2016. En este caso, el objetivo del autor fue predecir los valores de cierre del S&P 500, Dow Jones Industrial Average y Nasdaq 100 utilizando una red neuronal multicapa. En este caso, hablamos de un problema de regresión.

El modelo elegido para la elaboración de este estudio es Multilayer Perceptron feed-forward, correspondiente a la variedad ARX, para el que se configuraría una estructura final de 4 *layers* o capas. La literatura afirma que MLP tiene una gran capacidad para aproximar funciones y predecir valores futuros, por lo que será de interés para el autor comprobar si esto puede ser aplicado a la predicción del mercado.

Los *inputs* o variables utilizadas para la elaboración del modelo corresponden a los últimos 10 valores de cierre de cada índice junto con los 10 últimos valores de cierre de la prima (Fondos Federales) durante el período comprendido entre el 1 de enero de 1900 y el 1 de enero de 2010.

Para la evaluación de la precisión del modelo, el autor optó por utilizar el Coeficiente de Correlación de Pearson y la raíz del error cuadrático medio (*root mean squared errors*) como medidas.

Los resultados del estudio mostraron que la habilidad del modelo creado mediante MLP fue bastante buena a la hora de aproximar valores futuros para el caso del S&P 500. Según el autor, si después de entrenar el algoritmo este hubiera sido utilizado para tomar decisiones de inversión al final de cada día, habría sido posible superar el rendimiento del S&P 500 más de un 60% durante casi 4 meses

Sin embargo, resulta necesario comentar que los resultados de los otros dos índices fueron muy diferentes. Si se hubiera utilizado el algoritmo para tomar decisiones de inversión en el NASDAQ y Dow Jones, se habría obtenido un beneficio un 59% y 81% menor en comparación con el índice de referencia correspondiente. A pesar de que los resultados del coeficiente de correlación de Pearson y de la raíz del error cuadrático medio son muy similares a los obtenidos en el caso del S&P 500, las predicciones realizadas para el NASDAQ y Dow Jones son significativamente menos precisas. Esto nos indica que las medidas de error estándar utilizadas para evaluar la precisión del modelo no son apropiadas o son, por lo menos, insuficientes.

Continuando con su investigación, y en vista de los resultados de su modelo, el autor plantea la posibilidad de que se obtengan mejores resultados cuando el objetivo del estudio es predecir la dirección del mercado y no la cantidad de dicha variación, utilizando también otros métodos de evaluación. En este sentido, hace referencia a la consideración del problema como problema de clasificación en lugar de regresión y a la utilización del *hit ratio* como medida de evaluación, representando este las veces que el modelo acierta en predecir la dirección del mercado.

En definitiva, el autor propone a futuros investigadores considerar la predicción de precios como un problema de clasificación y no de regresión y utilizar el *hit ratio* como indicador de la precisión del modelo junto con más herramientas de evaluación. Sin embargo, a pesar de que afirma que las redes neuronales son capaces de aproximar funciones con cierta exactitud y que es posible obtener beneficios puntuales en el mercado con su utilización, considera que la obtención de mejores rendimientos a largo plazo es todavía dudosa y que los modelos elaborados con redes neuronales fallan a la hora de generalizar sus resultados (M. Harper, 2016)

Tabla 4. Modelo ANN 2

<b>Artificial Neural Networks - (M. Harper, 2016)</b>	
Clasificación del problema de predicción	Problema de regresión
Input	10 últimos precios de cierre de cada índice y 10 últimos valores de la prima
Modelo empleado	MLP
Output	S&P 500, Dow Jones Industrial Average y NASDAQ 100
Medidas de precision	<i>Root Mean Squared Errors</i> y <i>Coefficiente de Correlación de Pearson</i>
Resultados	Rendimiento 60% superior al del S&P500

Fuente: elaboración propia con datos de (M. Harper, 2016)

### **Estudio 3 – Qiu y Song (2016)**

#### *Modelo de predicción utilizando ANN y GA (Genetic Algorithms)*

El siguiente estudio que analizaremos fue elaborado por Qiu y Song en 2016. En este caso, los autores hicieron uso de un modelo combinado de ANN y GA (algoritmos genéticos) para llevar a cabo un estudio experimental con el objetivo de predecir de los movimientos del Nikkei 225 Index, el índice más conocido de la bolsa de valores de Tokio. Este índice está compuesto por 225 acciones igualmente ponderadas y ha sido calculado de manera diaria desde 1950, lo que nos permite disponer de una abundante cantidad de datos. Para la elaboración de este estudio, se utilizaron los datos recogidos entre enero de 2007 y diciembre de 2013, obteniendo un total de 1707 muestras (Qiu & Song, 2016). Para este estudio, se planteó el problema como un problema de clasificación.

En lo referente al modelo empleado, los autores optaron por un modelo basado en redes neuronales que sería entrenado mediante el algoritmo *Back Propagation*, uno de los algoritmos más utilizados en la predicción de series financieras. Sin embargo, debido a que la literatura demostraba determinadas limitaciones de este algoritmo a la hora de predecir los movimientos del mercado, los autores optaron por incorporar también otra clase de algoritmo al modelo, conocido como algoritmo genético (GA). La función del

algoritmo genético en este caso sería determinar el conjunto de pesos y *bias* óptimo para mejorar la precisión del modelo. De esta forma, el modelo finalmente creado, que pasaría a ser híbrido, podría evitar los problemas y limitaciones asociados al algoritmo BP.

En lo referente a las variables para el entrenamiento del modelo, estas se correspondían con una serie de indicadores técnicos agrupados en dos tipos (ver tablas 7 y 8); en este caso, se escogerían 13 indicadores del primer tipo y 9 del segundo para la elaboración del modelo. Finalmente, para la evaluación de la precisión del modelo se utilizaría el *hit ratio*, que indica el porcentaje de ensayos en los que la predicción del modelo fue correcta (Qiu & Song, 2016)

Los resultados obtenidos indicaron que, si se utilizan los indicadores del primer tipo como input, el *hit ratio* resultante alcanza el valor del 60.87%, lo que quiere decir que el modelo acierta en su predicción el 60.87% de las veces. Si, por el contrario, se utilizan los indicadores del segundo tipo, el *hit ratio* resultante alcanza el valor del 81.2%.

De estos resultados, los autores extrajeron dos conclusiones fundamentales: por un lado, es posible predecir, en gran medida, los movimientos del mercado y, por otro lado, el uso de los indicadores técnicos recogidos en el tipo 2 permite obtener mejores resultados de predicción.

En la segunda parte del estudio, los autores decidieron comparar los resultados de su modelo con los resultados que otros autores habían obtenido empleando algoritmos diferentes. Tal y como muestra la siguiente tabla, los resultados de la comparativa exhibieron un mayor *hit ratio* para el modelo GA-ANN, elaborado por Qiu y Song.

Tabla 5. Comparación modelos. Estudio ANN 3

Estudios	Métodos	Mercado	Hit Ratio (%)
Kim and Han	GA	Corea	61.70
Leung et al.	Modelo de clasificación	US, UK, Japón	68
Huang et al.	SVM	Japón	75
Kara et al.	BPNN	Estambul	75.74
<b>Qiu y Song</b>	GA-ANN	Japón	81.27

Fuente: elaboración propia con datos de (Qiu & Song, 2016)

En vista de los resultados obtenidos, los autores atribuyeron el éxito de su estudio a los indicadores utilizados y a la incorporación del algoritmo genético (GA) para la optimización del modelo. Tal y como afirman los autores, los errores de predicción presentes en otros estudios (como puede ser el caso de los mencionados previamente en este trabajo) pueden deberse a errores de planteamiento del problema, errores a la hora de elaborar el modelo correcto para dicho problema y errores de elección de las variables utilizadas para el entrenamiento.

Las conclusiones resultantes de este estudio son especialmente relevantes para este trabajo, ya que no solamente nos permiten aumentar los argumentos en contra de la hipótesis del mercado eficiente, sino que también nos aportan información sobre cuáles son los inputs más indicados a utilizar para la predicción del mercado de valores y añaden la posibilidad de combinar varios algoritmos para mejorar la precisión de los modelos (Qiu & Song, 2016)

Tabla 6. Modelo ANN 3

Artificial Neural Networks - (Qiu & Song, 2016)	
Clasificación del problema de predicción	Problema de clasificación
Input	Indicadores técnicos (ver tablas 7 y 8)
Modelo empleado	Back Propagation (ANN) + Genetic Algorithms (GA)
Output	Nikkei 225 Index
Medidas de precision	<i>Hit ratio</i>
Resultados	Ratio de acierto ( <i>hit ratio</i> ) del 81.27%

Fuente: elaboración propia con datos de (Qiu & Song, 2016)



Tabla 7. Indicadores técnicos tipo 1. Modelo ANN3

<b>Indicadores Técnicos Tipo 1</b>
Stochastic %K
Stochastic %D
Stochastic slow %D
Momentum
ROC (ratio de cambio)
Larry William's %R
A/O Oscillator (accumulation/distribution oscillator)
Disparidad en 5 días
Disparidad en 10 días
OSCP (oscilador del precio)
CCI (Commodity channel index)
RSI (Relative Strenght Index)

Fuente: elaboración propia con datos de (Qiu & Song, 2016).

Tabla 8. Indicadores técnicos tipo 2. Modelo ANN 3

<b>Indicadores Técnicos Tipo 2</b>
OBV
MA
BIAS
PSY
ASY <sub>5</sub>
ASY <sub>4</sub>
ASY <sub>3</sub>
ASY <sub>2</sub>
ASY <sub>1</sub>

Fuente: elaboración propia con datos de (Qiu & Song, 2016) – ver fórmulas en el Anexo II

## 5.2. Support Vector Machines

### **Estudio 4 – Saahil Madge (2015)**

#### Modelo de predicción utilizando el algoritmo SVM

En el siguiente estudio, elaborado por Madge en 2015, se utiliza un modelo creado usando SVM para estudiar los precios de cierre diarios de 34 empresas tecnológicas con el fin de calcular la volatilidad del precio y el *momentum* tanto de las acciones individuales como del sector en su conjunto. El objetivo principal de la investigación es predecir si el precio futuro será mayor o menor de lo que es actualmente, es decir, la dirección del mercado. Teniendo esto en consideración podemos afirmar que el problema se plantea como un problema de clasificación.

Los datos utilizados en este caso corresponden al precio de cierre de 34 compañías tecnológicas recogidas en el NASDAQ-100 Technology Index entre el año 2007 y el final del año 2014.

En lo referente al modelo SVM, se utiliza la función *radial kernel* para su elaboración, una de las funciones *kernel* más utilizadas por su capacidad de manejar diversas series de inputs. Las variables utilizadas para el entrenamiento del modelo fueron: volatilidad de la acción, *momentum* del precio, volatilidad del sector y *momentum* del sector. Con el fin de analizar diversas situaciones, se utilizó el modelo para realizar predicciones en el corto y en el largo plazo. Para la evaluación de la precisión del modelo se calculó el parámetro *accuracy*, concretamente, *mean prediction accuracy*.

Los resultados muestran que el modelo no es capaz de predecir correctamente la dirección del precio del día siguiente. En este sentido, el autor afirma que la utilización de SVM para la predicción del movimiento de los precios muestra ser similar a la probabilidad de predecir estos movimientos mediante el lanzamiento de una moneda (Madge, 2015), lo que vendría a reforzar las hipótesis del mercado eficiente y paseo aleatorio previamente mencionadas. Sin embargo, en el largo plazo, se encontraron evidencias que permiten afirmar un grado de predicción de hasta el 80% para algunas acciones. A pesar de lo positivo de este resultado, en la práctica resulta imposible adivinar en qué acciones

acertará el modelo, por lo que puede considerarse que los resultados del estudio refuerzan la hipótesis del mercado eficiente.

De los resultados del estudio el autor extrae varias conclusiones. En primer lugar, a pesar de que sus resultados reafirman la hipótesis del mercado eficiente, considera que, introduciendo las variables de entrenamiento adecuadas, el modelo sí podría ser capaz de sacar ventaja de las tendencias del mercado. En segundo lugar, afirma que es más conveniente utilizar como input el *momentum* de un stock para las predicciones a corto plazo y utilizar las tendencias del sector para las predicciones enfocadas al largo plazo. Por último, también hace una recomendación a futuros investigadores para que estos utilicen otras variables como inputs y otras ventanas temporales (Madge, 2015)

Tabla 9. Modelo SVM 1

<b>Support Vector Machines - Madge (2015)</b>	
Clasificación del problema de predicción	Clasificación
Input	<i>Momentum</i> de la acción y el mercado Volatilidad de la acción y del mercado
Modelo empleado	SVM – función <i>radial kernel</i>
Output	NASDAQ – 100
Medidas de precision	<i>mean prediction accuracy</i>
Resultados	El modelo no es capaz de predecir correctamente la dirección del precio de la acción y ni del mercado

Fuente: elaboración propia con datos de (Madge, 2015)

### **Estudio 5 – Di (2014)**

#### *Modelo de predicción utilizando el algoritmo SVM + Extremely Randomized Tree*

El siguiente estudio por evaluar, llevado a cabo por Di en 2014, se centra en la predicción de los movimientos de tres acciones concretas mediante la utilización de algoritmos SVM. Las tres compañías seleccionadas fueron Apple, Microsoft y Amazon, y el período escogido comprendía desde el 4 de enero de 2010 al 10 de diciembre de 2014.

Las variables de entrenamiento utilizadas fueron los siguientes indicadores técnicos: “Williams %R”, “Rate of Change”, “Momentum”, “Relative Strength Index”, “Commodity Channel Index”, “Average Directional Index”, “Triple Exponential Moving Average”, “Moving Average Convergence Divergence”, “On Balance Volume”, “Times Series Forecasting”, “Average True Range”, “Money Flow Index”.

Debido al gran número de elementos y el consecuente ruido que estos podrían llegar a generar, el autor decidió utilizar el algoritmo Extremely Randomized Tree para que este eligiese solo aquellos indicadores más relevantes para cada acción. Por otro lado, se decidió probar el modelo en cuatro ventanas temporales distintas, analizando las predicciones de los 3, 5, 7 y 10 días siguientes. Los resultados del estudio pueden observarse en la tabla 10.

Tabla 10. Comparativa de la precisión. Modelo SVM 2

Compañía/ <i>accuracy</i>	Sig. 3 días	Sig. 5 días	Sig. 7 días	Sig. 10 días
Apple	73.4%	71.41%	70.25%	71.13%
Amazon	63%	65%	61.5%	71.25%
Microsoft	64.5%	73%	77.125%	77.25%

Fuente: elaboración propia con datos de (Di, 2014)

Tabla 11. Modelo SVM 2

Support Vector Machines - (Di, 2014)	
Clasificación del problema de predicción	Clasificación
Input	<b>Indicadores:</b> WILLR, ROCR, MOM, RSI, CCI, ADX, TRIX, MACD, OBV, TSF, ATR, MFI
Modelo empleado	SVM + Extremely Randomized Tree
Output	Apple, Microsoft, Amazon
Medida de precisión	<i>Accuracy</i> (exactitud)
Resultados	Predicción de entre el 70 y el 73.4% utilizando indicadores generalistas

Fuente: elaboración propia con datos de (Di, 2014)

En los datos resultantes puede observarse que, las mejores predicciones para el corto plazo se han realizado para el caso de la compañía Apple. Según el autor, esta diferencia podría atribuirse a las variables utilizadas en el modelo: en el caso de Apple los indicadores más generalistas tenían un peso mayor frente a otros, lo que reducía significativamente la varianza. Además, afirma que esta información resulta muy relevante de cara a investigaciones futuras, pues indica qué clase de indicadores podría ser más adecuado para entrenar el modelo (Di, 2014)

### **Estudio 6 – Yu, Wang y Keung Lai (2005)**

#### *Modelo de predicción utilizando el modelo GA-SVM*

El siguiente estudio, elaborado por Yu, Wang y Keung Lai en 2005, buscaba elaborar un modelo utilizando la combinación SVM-GA (Support Vector Machines y Algoritmo Genético) para predecir la dirección del mercado. Se trata por tanto de un problema de clasificación.

Anteriormente se explicaba la capacidad de los algoritmos SVM de evitar el problema de mínimos locales relacionado con las redes neuronales artificiales, motivo por el que se consideraba que este tipo de algoritmos podría tener mejor capacidad de generalización y ser más recomendable para la solución de problemas de predicción. En este caso, los autores optan por dar un paso más incorporando el algoritmo genético (GA) al modelo SVM con el fin de hacer este menos complejo.

Los datos utilizados para la elaboración de este estudio corresponden al precio de cierre del índice S&P 500 entre el 1 de enero del 2000 y el 31 de diciembre de 2004. Gracias a ellos, se elaboraron 18 indicadores técnicos que servirían para entrenar el modelo.

Para llevar a cabo la evaluación de los resultados del modelo, estos se compararían con los proporcionados por un modelo estadístico (ARIMA), un modelo elaborado con redes neuronales artificiales back-propagation (BPNN), otro modelo SVM sin incorporar el algoritmo genético y el modelo *paseo aleatorio* o Random Walk (RW).

En el caso del modelo GA-SVM, la selección de variables se llevaría a cabo mediante el uso de un algoritmo genético, de manera que sería este algoritmo el encargado de elegir

qué variables servirían como *inputs* para el entrenamiento del modelo. De los 18 indicadores técnicos que se habían seleccionado, el algoritmo genético finalmente eligió 11, siendo estos: price (P), stochastic oscillator (SO), rate of change (ROC), moving average (MA), moving variance ratio (MVR), moving average convergence & divergence (MACD), disparity5 (D5), price oscillator (OSCP), commodity channel index (CCI), relative strength index (RSI) y linear regression line (LRL). En lo referente a los resultados obtenidos, tal y como muestra la siguiente tabla (12) pudieron apreciarse diferencias significativas entre los modelos.

Tabla 12. Comparación modelos. Modelo SVM 3

Modelo	Hit ratio (%)
RW	51.06
ARIMA	56.13
BPNN	69.78
SVM	78.65
GA-SVM	84.57

Fuente: elaboración propia con datos de (Yu, Wang, & Keung Lai, 2005)

Como puede comprobarse, el modelo que obtuvo un mayor hit ratio fue el GASVM con un 84.57%. Teniendo en cuenta estos resultados empíricos, los autores afirman que, aunque el modelo SVM alcanza unos niveles de precisión aceptables, la combinación con el algoritmo genético mejora sustancialmente la capacidad de predicción del modelo (Yu, Wang, & Keung Lai, 2005)

Tabla 13. Modelo SVM 3

Random Forest - (Yu, Wang, & Keung Lai, 2005)	
Clasificación del problema de predicción	Clasificación
Input	Indicadores técnicos: SO, ROC, MA, MVR, MACD, D5, OSCP, CCI, RSI y LRL
Modelo empleado	Support Vector Machines + Algoritmo Genético
Resultados	Precisión del 84.57%
Output	S&P 500
Medidas de precisión	Hit ratio

Fuente: elaboración propia con datos de (Yu, Wang, & Keung Lai, 2005)

### 5.3. Random Forest

#### **Estudio 7 - Khaidem, Saha y Roy Dey (2016)**

##### *Modelo de predicción utilizando el algoritmo Random Forest*

El primer estudio que analizaremos dentro de este apartado fue llevado a cabo por Khaidem, Saha y Roy Dey en 2016. El estudio se centra en la elaboración de un modelo de predicción utilizando el algoritmo *Random Forest* y está dividido en dos partes. En primer lugar, se realiza una prueba de predicción utilizando la información disponible de tres compañías cotizadas. En segundo lugar, se comparan los resultados obtenidos con otros estudios similares. En este caso, el problema de predicción será considerado de clasificación y no de regresión.

Los datos elegidos para el entrenamiento del algoritmo corresponden al precio de cierre de las compañías Apple, General Electric y Samsung. Dada la similitud de resultados obtenidos y con el fin de simplificar la presentación de este estudio, únicamente se mostrarán los resultados obtenidos del análisis de la compañía Apple. Previamente a la introducción de los datos en el modelo, las series temporales obtenidas fueron sometidas a un proceso de suavizamiento exponencial, también conocido como *exponential smoothing*, para dotar de un peso mayor a los valores más recientes frente a los más antiguos. Realizado este proceso, los autores configurarían una serie de indicadores técnicos con los datos resultantes, que servirían como variables de entrenamiento para el modelo. Los indicadores técnicos utilizados fueron: *Relative Strength Index (RSI)*, *Stochastic Oscillator*, *Williams %R*, *MACD*, *Price Rate of Change* y *On Balance Volume*.

Para la elaboración de este modelo se utilizaron 30 árboles de decisión, y se decidió utilizar una serie de parámetros junto con la representación gráfica de la curva ROC (*Receiver Operating Characteristic*) para evaluar su desempeño. Los parámetros utilizados en este caso fueron: precisión (*precisión*), exactitud (*accuracy*), sensibilidad (*sensitivity*) y especificidad (*specificity*). Las fórmulas de dichos parámetros pueden encontrarse en el Anexo I.

Los resultados obtenidos indicaron que, de los 30 árboles utilizados, 29 predijeron una subida del precio de la acción mientras que solo uno predijo que el valor de la acción bajaría. Así mismo, los investigadores pudieron observar que, a medida que el número de

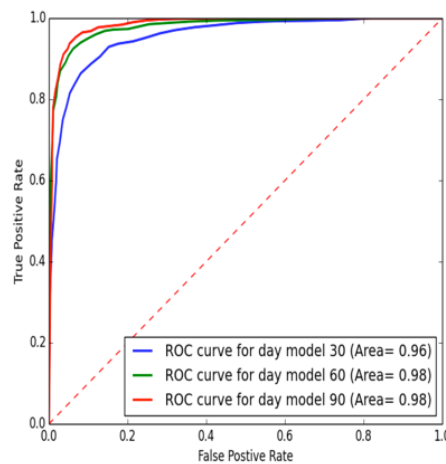
días de entrenamiento y el número de árboles aumentaba, el porcentaje de error de predicción del modelo disminuía.

Tabla 14. Resultados para las acciones de Apple Inc. Modelo RF 1

Ventana temporal de entrenamiento	% de exactitud ( <i>accuracy</i> )	Precisión (precision)	Sensibilidad ( <i>recall</i> )	Especificidad ( <i>specificity</i> )
1 mes	88.264	0.89263	0.90724	0.84848
2 meses	93.065	0.94154	0.93858	0.91973
3 meses	94.533	0.94548	0.96120	0.92341

Fuente: elaboración propia con datos de (Khaidem, Saha, & Roy Dey, 2016)

Ilustración 6. Curva ROC Apple Inc. Modelo RF 1



Fuente: (Khaidem, Saha, & Roy Dey, 2016)

En la segunda parte del estudio se comparan los resultados obtenidos con otros estudios en los que se utilizó un modelo diferente. Concretamente, se revisaron los resultados que otros autores habían obtenido tras utilizar el algoritmo SVM, previamente explicado en este trabajo.

Los resultados de la comparativa con el estudio de Dai y Zhang (2013), revelaron que, a pesar de que el modelo SVM presentaba un porcentaje de acierto del 79,3% cuando la ventana temporal era de 44 días, la utilización del algoritmo *random forest* incrementaba esta precisión hasta el 96,92% con una ventana temporal de 88 días utilizando los mismos datos.



En la segunda comparativa, en la que se analizó el estudio realizado por Xinjie en 2014, los autores pudieron observar un mayor porcentaje de precisión cuando el algoritmo utilizado era *random forest* en lugar de SVM. Al comparar los datos de la misma empresa, en las mismas fechas, utilizando el mismo horizonte temporal (movimientos del mercado los 3, 5, 7 y 10 días siguientes) pudo observarse un incremento importante de la precisión del modelo cuando el algoritmo utilizado para la predicción se trataba de *random forest*. Por otro lado, las predicciones realizadas con el algoritmo SVM demostraron ir perdiendo precisión a medida que los días aumentaban. La siguiente tabla muestra los resultados de dicha comparativa.

Tabla 15. Comparación modelos. Modelo RF 1

\*período comprendido entre el 4 de enero de 2010 y el 10 de diciembre de 2014

Algoritmo utilizado/ accuracy	3 días siguientes	5 días siguientes	7 días siguientes	10 días siguientes
SVM	73.4%	71.41%	70.25%	71.13%
Random Forest	85.197%	83.88%	88.11%	92.08%

Fuente: elaboración propia con datos de (Khaidem, Saha, & Roy Dey, 2016)

Una vez finalizadas las comparativas y analizados los resultados, los autores proceden a exponer las conclusiones de su trabajo. En primer lugar, consideran que, a pesar de que el modelo SVM tiene un porcentaje de precisión considerablemente alto, la utilización de *random forest* para problemas de clasificación es todavía mejor, y atribuyen este mejor desempeño a dos factores: el mejor procesamiento de los datos gracias a *exponential smoothing* y el aumento de la ventana temporal en la que el modelo es entrenado (Khaidem, Saha, & Roy Dey, 2016).

Tabla 16. Modelo RF 1

Random Forest - (Khaidem, Saha, & Roy Dey, 2016)	
Clasificación del problema de predicción	Problema de clasificación
Input	Indicadores técnicos Tabla 17
Modelo empleado	Random Forest + <i>exponential smoothing</i>
Output	Apple, General Electric, Samsung
Medidas de precisión	Curvas ROC y <i>accuracy</i>
Resultados	% de precisión de entre el 85-95%

Fuente: elaboración propia con datos de (Khaidem, Saha, & Roy Dey, 2016)

Tabla 17. Indicadores Técnicos. Modelo RF 1

Indicadores técnicos
Relative Strength Index
Stochastic Oscillator
Williams %R
Moving Average Convergence Divergence
Price Rate of Change
On Balance Volume

Fuente: elaboración propia con datos de (Khaidem, Saha, & Roy Dey, 2016)

### **Estudio 8 - (Bafandeh Inmandoust & Bolandraftar, 2014)**

#### *Modelo de predicción basado en el algoritmo Random Forest*

El segundo estudio que se analizará dentro de esta categoría corresponde a los autores Bafandeh Imandoust y Bolandraftar. El objetivo principal de este trabajo de investigación corresponde a la elaboración de tres modelos de predicción, basados en tres técnicas de clasificación diferentes, para la predicción del mercado de valores de Teherán. Los modelos utilizados son: *Random Forest*, *Decision Tree* y *Naïve Bayesian Classifier*. Como previamente se anticipaba, el problema es planteado como un problema de clasificación.

Los datos utilizados para la elaboración de este trabajo corresponden al precio de cierre del índice TSE durante el período comprendido entre el 17 de abril del año 2007 y el 18 de marzo del año 2012. En cuanto a los inputs utilizados para el entrenamiento de los modelos, se optó por hacer tres grupos: el primero basado únicamente en indicadores técnicos, el segundo en indicadores fundamentales y el tercero combinando indicadores técnicos y fundamentales. Con el fin de comprobar cuál de las clases de indicadores es más adecuada para la elaboración del modelo, se realizó una comparativa entre los grupos, elaborando los modelos en tres partes según el tipo de input utilizado.

Los indicadores técnicos utilizados están recogidos en la tabla 18. Para el caso de los indicadores fundamentales, se utilizarían los precios del petróleo y el oro, así como la cotización USD/IRR.

Tabla 18. Indicadores técnicos modelo RF 2.

Nombre del indicador
Simple 10-day moving average
Weighted 10-day moving average
Momentum
Stochastick K%
Stochastic D%
RSI (Relative Strength Index)
MACD (Moving Average Convergence Divergence)
Larry William's R%
A/D (Accumulation/Distribution) Oscillator
CCI (Commodity Channel Index)

Fuente: elaboración propia con datos de (Bafandeh Inmandoust & Bolandraftar, 2014)

Los resultados obtenidos en el caso de *Random Forest* muestran un porcentaje de acierto del 78.81% cuando los inputs utilizados para entrenar el modelo son únicamente indicadores técnicos. Para el caso del uso único de indicadores fundamentales, el porcentaje de acierto desciende al 61.86%. Por último, el resultado de combinar indicadores técnicos y fundamentales nos muestra un porcentaje de acierto del 78.39%, lo que quiere decir que el modelo predijo correctamente la dirección del mercado un 78.93% de las veces.

Sin embargo, cabría mencionar los resultados obtenidos del modelo *Decision Tree*, pues, en este caso, el porcentaje de acierto alcanzó el valor del 80.08% cuando se emplearon indicadores técnicos e indicadores técnico-fundamentales como inputs.

Como conclusión de este estudio, los autores están de acuerdo en afirmar que es posible predecir en gran medida la dirección del mercado de valores de Teherán utilizando técnicas de *data mining*. Por otro lado, consideran que los indicadores técnicos son más adecuados para plantear este tipo de problemas de clasificación, ya que con su uso se ha obtenido un porcentaje de acierto de hasta el 80.08% para el modelo realizado con *Decision Tree* y del 78.81% en el caso de *Random Forest*.

Tabla 19. Modelo RF 2

<b>Random Forest - (Bafandeh Inmandoust &amp; Bolandraftar, 2014)</b>	
Clasificación del problema de predicción	Problema de clasificación
Input	Indicadores técnicos y fundamentales
Modelo empleado	Random Forest
Output	Mercado de valores de Teherán
Medidas de precisión	<i>accuracy</i>
Resultados	78.81% de acierto con el modelo Random Forest

Fuente: elaboración propia con datos de (Bafandeh Inmandoust & Bolandraftar, 2014)

### **Estudio 9 – Basak, Saha, Kar, Khaidem y Dey (2019)**

#### Modelo de predicción basado en el algoritmo Random Forest

En este estudio, el objetivo principal de los autores consistía en el diseño de un modelo capaz de aprender de la información del mercado, gracias a *machine learning*, para predecir la dirección en la que los precios se moverían a futuro. En este caso, el problema se plantea como un problema de clasificación y no de regresión.

Al comienzo de este estudio, los autores hacen un repaso de la literatura previa en materia de predicción de precios mediante *machine learning*, parte de ella recogida previamente en este trabajo, y llegan a la conclusión de que apenas se han utilizado los algoritmos de aprendizaje *tree-ensemble* con esta finalidad. Por este motivo, toman la decisión de basar su estudio en modelos que utilizan árboles de decisión, siendo estos: *Random Forest (RF)* y *Gradient Boosted Trees (GBDT)*. En este caso, nos centraremos únicamente en el análisis de los resultados obtenidos por el modelo *Random Forest*.

Los datos utilizados para elaborar el modelo corresponden a diez compañías, seleccionadas de forma aleatoria, de las cuales se estudió el precio de cierre y el volumen entre otras variables para elaborar una serie de indicadores técnicos que servirían como variables de entrenamiento para el modelo. El período de tiempo del cual se recogieron los datos se extiende desde el momento en el que las acciones entraron en el mercado respectivamente hasta el 3 de febrero de 2017.

Concretamente, se utilizarían los siguientes indicadores técnicos como inputs: *Relative Strength Index (RSI)*, *Stochastic Oscillator*, *Williams Percentage Range*, *Moving Average Convergence Divergence (MACD)*, *Price Rate of Change (PROC)* y *On Balance Volume (OBV)*.

Una de las características novedosas de este estudio reside en la utilización del método de suavizamiento exponencial, también conocido como *exponential smoothing*, que elimina variables aleatorias de los datos y hace el proceso de aprendizaje más sencillo. Este método se emplearía como paso posterior a la recolección de los datos (precio de cierre, volumen etc. de cada una de las acciones seleccionadas). Una vez realizado este proceso se extraerían los indicadores técnicos para entrenar al modelo. En este sentido, es igualmente importante mencionar que los datos más recientes tendrían un peso mayor frente a los datos más antiguos.

Otro de los puntos a destacar dentro de este estudio es la ventana temporal utilizada. A diferencia de estudios anteriores, cuyas ventanas temporales ascendían a 44 días (como pudimos observar previamente en el caso de Khaidem, Saha y Roy Dey (2016)), en este caso los autores optan por extender el período sobre el que el modelo haría predicciones hasta los 90 días.

En lo referente a la evaluación de los resultados del modelo, los autores han optado por la utilización de diversos parámetros, entre ellos: F-score, precisión (*precisión*), exactitud (*accuracy*), especificidad (*specificity*). Una vez construido el modelo, que finalmente constaría de tres árboles de decisión, se pasaría a analizar la precisión de este.

Los resultados de la clasificación realizada por el modelo *Random Forest* aportan información de gran relevancia, pues el modelo fue capaz de alcanzar una exactitud (*accuracy*) del 93.02% en el caso de Apple y del 93.99% en el caso de Toyota para una ventana temporal de 90 días. En esta misma línea, puede apreciarse como se produce un aumento de la precisión a medida que la ventana temporal de trading aumenta, lo que quiere decir que los algoritmos son capaces de aprender de los datos utilizados.

Teniendo en cuenta los resultados anteriores, los autores afirman que se da una mejora significativa de la precisión cuando la ventana temporal pasa de 3 o 5 días a ser de 60, 90

o 120 días, motivo por el que consideran fallaron estudios previos. Dado el éxito de su trabajo, los autores recomiendan el tratamiento previo de los datos con el método *exponential smoothing* y el uso de algoritmos capaces de trabajar en el largo plazo para poder así aumentar la ventana temporal y mejorar la precisión del modelo (Basak, Saha, Kar, Khaidem, & Dey, 2019).

Tabla 20. Modelo RF 3

Random Forest - (Basak, Saha, Kar, Khaidem, & Dey, 2019)	
Input	Indicadores técnicos: RSI, SO, WPR, MACD, PROC, OBV
Modelo empleado	Random Forest + <i>exponential smoothing</i>
Resultados	% de precisión de hasta un 93.99%
Output	Apple, Toyota,
Medidas de precisión	F-score, precisión, exactitud y especificidad

Fuente: elaboración propia con datos de (Basak, Saha, Kar, Khaidem, & Dey, 2019)

## 6. CONCLUSIONES

Una vez revisada parte de la literatura y los estudios relacionados con la utilización de la inteligencia artificial como herramienta de inversión, podemos extraer una serie de conclusiones que nos servirán para dar respuesta a las preguntas planteadas al inicio de este trabajo. Por otro lado, con la formulación de estas conclusiones se pretende presentar una serie de recomendaciones de cara a investigaciones futuras.

En primer lugar, haciendo referencia al propósito principal de esta investigación, nos gustaría destacar que existen estudios y datos empíricos suficientes como para considerar la inteligencia artificial una interesante herramienta para la inversión. Estas afirmaciones no implican que siempre vaya a ser posible obtener un beneficio superior a la media mediante la utilización de inteligencia artificial, sin embargo, sí nos es posible afirmar, gracias a los estudios analizados, que el uso de modelos de *machine learning* puede derivar en la elaboración de una estrategia de inversión que genere buenos rendimientos.

En lo que a los objetivos secundarios de la investigación se refiere, hemos optado por desglosar las conclusiones relacionadas con ellos en sub-apartados.

### *Hipótesis del mercado eficiente y paseo aleatorio*

En primer lugar, en lo que a las teorías del mercado eficiente y paseo aleatorio se refiere, creemos que no hay datos suficientes como para rechazar por completo ambas teorías. Si bien es cierto que los estudios analizados nos muestran que es posible predecir en mayor o menor medida los movimientos del mercado de valores, no puede considerarse que los modelos elaborados con inteligencia artificial sean capaces de predecir con exactitud los futuros movimientos del mercado.

Cabe destacar nuevamente que estas afirmaciones no implican la inutilidad de las herramientas de inteligencia artificial en la elaboración de estrategias de inversión, pues se ha podido observar que son capaces de predecir los movimientos del mercado en un alto porcentaje, por lo que tendrían cabida en la elaboración de una estrategia de inversión. Sin embargo, la diversidad de resultados obtenidos no nos permite rechazar con rotundidad estas teorías.

### *Elección del modelo adecuado para elaborar la estrategia de inversión.*

En referencia a cuál es el mejor modelo para la elaboración de una estrategia de inversión en función a su capacidad de predicción del mercado, una vez realizado nuestro estudio podemos considerar que se trata de un tema que requiere de mayor investigación. Los motivos que nos llevan a tal conclusión guardan relación con la disparidad de los resultados obtenidos. Sin embargo, cabe mencionar que ha sido posible la extracción de una serie de conclusiones referentes a la elaboración del modelo que pueden resultar de gran interés de cara a futuras investigaciones.

En este aspecto, debe exponerse en primer lugar que se ha podido observar un aumento de la precisión del modelo cuando el problema de predicción es considerado de clasificación y no de regresión. Así lo comentaban previamente otros estudios como el elaborado por M. Harper (2016) y ha podido reafirmarse a lo largo del presente trabajo.

En relación con los inputs utilizados para la elaboración del modelo, una vez revisados diversos estudios empíricos, podemos afirmar que la elección de las variables de entrenamiento supone un punto clave en el éxito del modelo y que de ello dependen en gran medida los resultados obtenidos por este. Así, autores como Madge (2015) afirman que sería conveniente avanzar en la investigación sobre cuáles son los inputs más adecuados para llevar a cabo este tipo de problemas de predicción. De hecho, a lo largo de este trabajo han podido observarse diferencias significativas entre los estudios en los

que las variables se eligieron arbitrariamente frente a aquellos en los que las variables fueron elegidas utilizando otros métodos.

En este sentido, nuestra investigación nos ha permitido conocer que el uso de indicadores técnicos como variables de entrenamiento parece aportar mejores resultados. De acuerdo con el trabajo de Qiu y Song (2016) dentro del tipo de indicadores técnicos a utilizar, cabría destacar los recogidos en la Tabla 8, pues estos demostraron proporcionar mejores resultados que otros indicadores. Este punto puede ser muy interesante de cara a futuras investigaciones, de hecho, los autores están de acuerdo en afirmar que sería conveniente continuar estudiando el uso de los indicadores técnicos recogidos en la Tabla 8 como variables de entrenamiento.

Otro apunte interesante que podemos extraer de cara a la elaboración del modelo de predicción es la posibilidad de combinar distintos algoritmos para conseguir mejores resultados. Como hemos podido analizar a lo largo de este trabajo, cada algoritmo tiene sus debilidades y fortalezas; si se trabajase conjuntamente con varios algoritmos podría darse la situación en la que las debilidades de uno se vieran minimizadas por las fortalezas de otro, mejorando así la predicción del modelo (Kumar & Thenmozhi, 2006). Esta situación ha podido observarse cuando se han incluido algoritmos genéticos (GA); de hecho, la utilización de este tipo de algoritmos para la optimización del modelo ha resultado en predicciones más precisas tanto en el caso de las redes neuronales (Qiu & Song, 2016) como en *Support Vector Machines* (Yu, Wang, & Keung Lai, 2005).

En definitiva, en lo que a la elaboración del modelo se refiere, puede considerarse que tanto la optimización de las variables de entrenamiento como la categorización del problema como *problema de clasificación* resultan elementos altamente influyentes en la precisión del modelo.

Como aclaración final sobre la información recogida en este trabajo, podemos considerar que el uso de inteligencia artificial en la inversión, aunque aparentemente complicado, es un ámbito de estudio que merece nuestra atención por las posibilidades que puede brindarnos. Precisamente con el objetivo de reducir esa complejidad se requiere de la elaboración de más estudios que sigan la línea del presente trabajo, pues sólo así se conseguirá una mejor comprensión de estas nuevas herramientas y se podrá hacer un uso eficiente de las mismas.



## 7. BIBLIOGRAFÍA

- Abu-Mostafa, Y. S., & Atiya, A. F. (1996). Introduction to Financial Forecasting. *Applied Intelligence*, 6, 205-213.
- Bafandeh Inmandoust, S., & Bolandraftar, M. (junio de 2014). Forecasting the direction of stock market index movement using three data mining techniques: the case of Tehran Stock Exchange. *International Journal of Engineering Research and Applications*, 4(6), 106-117.
- Bagnato, J. I. (12 de diciembre de 2017). Qué es overfitting y underfitting y cómo solucionarlo. Obtenido de *Aprende Machine Learning*: <https://www.aprendemachinlearning.com/que-es-overfitting-y-underfitting-y-como-solucionarlo/>
- Basak, S., Saha, S., Kar, S., Khaidem, L., & Dey, S. R. (2019). Predicting the Direction of Stock Market Price Using Tree Based Classifiers. *The North American Journal of Economics and Finance*, 552-567.
- Bonnie G. Buchanan, P. F. (2019). *Artificial intelligence in finance*. Londres: The Alan Turing Institute.
- Cai, E. (26 de mayo de 2014). Machine Learning and Applied Statistics Lesson of the Day – Sensitivity and Specificity. Obtenido de *The Chemical Statistician*: <https://chemicalstatistician.wordpress.com/2014/05/26/machine-learning-and-applied-statistics-lesson-of-the-day-sensitivity-and-specificity/>
- Chen, J. (18 de abril de 2020). Relative Strenght Index (RSI). Obtenido de *Investopedia*: <https://www.investopedia.com/terms/r/rsi.asp>
- Chen, J. (24 de abril de 2019). Technical Indicator. Obtenido de *Investopedia*: <https://www.investopedia.com/terms/t/technicalindicator.asp>
- Cocinau, C.-L., Grigoryan, H., & Uscatu, C. (2017). STOCK MARKET FORECASTING USING WAVELET DENOISING TEHNIQUE AND SUPPORT VECTOR MACHINES. *Proceedings of the IE 2017 International Conference*. Bucarest: Departamento de Informática y Cibernética Económica de la Academia de Estudios Económicos de Bucarest.
- Cortes, C., & Vapnik, V. (1995). *Support-Vector Networks*. Kluwer Academic Publishers, 273-279.
- Crypthopper. (2020). Trading 101: 4 types of indicators you need to know. Obtenido de *Cyptohopper*: <https://www.crypthopper.com/blog/168-trading-101-technical-analysis-for-beginners>

- D. Arnott, R., Hsu, J., Kalesnik, V., & Tindall, P. (2013). The surprising Alpha from Malkiel's Monkey and Upside-Down Strategies. *The Journal of Portfolio Management*, 39(4).
- Das, S. R. (2019). *The Future of FinTech*. Santa Clara: Santa Clara University.
- Dharmaraja, S., Vineet, K., & Abhishek, M. (2019). Indian stock market prediction using artificial neural networks on tick data. Springer Open.
- Di, X. (2014). *Stock Trend Prediction with Technical Indicators using SVM*. Stanford: Leland Stanford Junior University.
- Fama, E. F. (Diciembre de 1991). Efficient Capital Markets: II. *The Journal of Finance*, 46(5), 1575-1617.
- Fernández-Rodríguez, F., González-Martel, C., & Sosvilla-Rivero, S. (2000). On the profitability of technical trading rules based on artificial neural networks: Evidence from the Madrid stock market. *Economics Letters*, 69, 89-94.
- Fernandez, H. (2019). Mercado de valores: Definición, características e importancia. Obtenido de Economía TIC: <https://economytic.com/mercado-de-valores/>
- Financial Stability Board. (2017). Artificial intelligence and machine learning in financial services. Market developments and financial stability implications. Financial Stability Board.
- Gautam, B., Gupta, D. S., Awasthi, D. S., & Gautam, S. (2019). Securities Analysis and Portfolio Management using Artificial Neural Networks. *International Journal of Research and Analytical Reviews*.
- Gharoie Ahangar, R., Yahyazadehfar, M., & Pournaghshband, H. (Febrero de 2010). The Comparison of Methods Artificial Neural Network with Linear Regression Using Specific Variables for Prediction Stock Price in Tehran Stock Exchange. (*IJCSIS*) *International Journal of Computer Science and Information Security*, 7(2), 38-46.
- Ghoneim, S. (2 de Abril de 2019). Accuracy, Recall, Precision, F-Score & Specificity, which to optimize on? Obtenido de Towards data science: <https://towardsdatascience.com/accuracy-recall-precision-f-score-specificity-which-to-optimize-on-867d3f11124>
- Gomber, P., Arndt, B., Lutat, M., & Uhle, T. (2011). *High-Frequency Trading*. Frankfurt: Goethe Universität Frankfurt am Main.
- Google Developers. (10 de febrero de 2020). Clasificación: Exactitud. Obtenido de Machine Learning: <https://developers.google.com/machine-learning/crash-course/classification/accuracy?hl=es-419>

- Google Developers. (10 de febrero de 2020). Clasificación: Umbral. Obtenido de Machine Learning Crash Course: <https://developers.google.com/machine-learning/crash-course/classification/thresholding>
- Google Developers. (10 de febrero de 2020). Classification: Precision and Recall. Obtenido de Machine Learning Crash Course: <https://developers.google.com/machine-learning/crash-course/classification/precision-and-recall>
- Google Developers. (10 de febrero de 2020). Classification: ROC Curve and AUC. Obtenido de Machine Learning Crash Course: <https://developers.google.com/machine-learning/crash-course/classification/roc-and-auc>
- Google Developers. (2 de febrero de 2020). Machine Learning Crash Course. Obtenido de Framing: Key ML Terminology: <https://developers.google.com/machine-learning/crash-course/framing/ml-terminology>
- Google Developers. (5 de marzo de 2020). Glosario sobre aprendizaje automático. Obtenido de Machine Learning Course: [https://developers.google.com/machine-learning/crash-course/glossary#top\\_of\\_page](https://developers.google.com/machine-learning/crash-course/glossary#top_of_page)
- Google Developers. (febrero de 2020.). Training and Test Sets: Splitting Data. Obtenido de Machine Learning Crash Course: <https://developers.google.com/machine-learning/crash-course/training-and-test-sets/splitting-data>
- Grosse, R. (2017). Lecture 9: Generalization. Intro to Neural Networks and Machine Learning. CSC 321 Winter 2018. Toronto: University of Toronto.
- Guresen, E., Kayakutlu, G., & Daim, T. U. (2011). Using artificial neural network models in stock market index prediction. *Expert System with Applications*.
- Hayes, A. (30 de junio de 2019). Stochastic Oscillator Definition. Obtenido de Investopedia: <https://www.investopedia.com/terms/s/stochasticoscillator.asp>
- Hegazy, O., Soliman, O. S., & Abdul Salam, M. (diciembre de 2013). A Machine Learning Model for Stock Market Prediction. *International Journal of Computer Science and Telecommunications*, 4(12), 17-23.
- Hu, Y., Liu, K., Zhang, X., Su, L., Ngai, E., & Liu, M. (noviembre de 2015). Application of evolutionary computation for rule discovery in stock algorithmic trading: A literature review. *Applied Soft Computing*, 36, 534-551.
- J. Murphy, J. (1999). *Technical Analysis of the Financial Markets. A comprehensive guide to trading methods and applications*. Paramus: New York Institute of Finance.

- Khaidem, L., Saha, S., & Roy Dey, S. (2016). Predicting the direction of stock market prices using random forest. *Applied Mathematical Finance*, 1-20.
- Kimoto, T., Asakawa, K., & Yoda, M. (1990). Stock market prediction system with modular neural networks. 1990 IJCNN international joint conference on neural networks, (págs. 1-6).
- Kumar, M., & Thenmozhi, M. (2006). FORECASTING STOCK INDEX MOVEMENT: A COMPARISION OF SUPPORT VECTOR MACHINES AND RANDOM FOREST. Indian Institute of Capital Markets 9th Capital Markets Conference Paper.
- Larrañaga, P., Inza, I., & Moujahid, A. (1997). *Redes Neuronales*. Universidad del País Vasco - Euskal Herriko Unibertsitatea.
- M. Harper, K. (2016). *Challenging the Efficient Market Hypothesis with Dynamically Trained Artificial Neural Networks*. Jacksonville: University of North Florida.
- Madge, S. (2015). *Predicting Stock Price Direction using Support Vector Machines*. Independent Work Report.
- Malkiel, B. G. (1973). *A Random Walk Down Wall Street*. New York: W. W. Norton & Company, Inc.
- Malkiel, B. G. (2003). The Efficient Market Hypothesis and Its Critics. *Journal of Economic Perspectives*, 17(1), 59-82.
- Martinez Heras, J. (2020). ¿Clasificación o Regresión? Obtenido de IArtificial.net: <https://iartificial.net/clasificacion-o-regresion/>
- Pakdaman Naeini, M., Taremian, H., & Baradaran Hashemi, H. (2010). Stock Market Value Prediction Using Neural Networks. 2010 International Conference on Computer Information Systems and Industrial Management Applications (CISIM), (págs. 132-136). Krakow, Poland.
- Peart, A. (2 de junio de 2017). Homage to John McCarthy, the Father of Artificial Intelligence (AI). Obtenido de Artificial Solutions: <https://www.artificial-solutions.com/blog/homage-to-john-mccarthy-the-father-of-artificial-intelligence>
- Qiu, M., & Song, Y. (2016). Predicting the Direction of Stock Market Index Movement Using an Optimized Artificial Neural Network Model. *PLOS ONE*, 11(5), e0155133.
- Sachin Sampat, P., Kailash, P., & Jain, M. (2016). Stock Market Prediction Using Support Vector Machine. *International Journal of Current Trends in Engineering & Technology*, 2(1), 18-25.

- Samek, D., & Varacha, P. (2013). Time Series Prediction Using Artificial Neural Networks: Influence of The Input Vector Size. *International Journal of Mathematical Models and Methods in Applied Sciences*, 38-46.
- Sampat Patil, M. S., Patidar, P. K., & Jain, P. M. (2016). Stock Market Prediction Using Support Vector Machine. *International Journal of Current Trends in Engineering & Technology*, 02(01), 18-25.
- Seth, S. S. (2019, junio 25). *The World of High-Frequency Algorithmic Trading*. Obtenido de Investopedia: <https://www.investopedia.com/articles/investing/091615/world-high-frequency-algorithmic-trading.asp#the-current-state-of-hft>
- Shah, D., Isah, H., & Zulkernine, F. (2019). Stock Market Analysis: A Review and Taxonomy of Prediction Techniques. *International Journal of Financial Studies*.
- Tan, Z., Yan, Z., & Zhu, G. (2019). Stock selection with random forest: An exploitation of excess return in the Chinese stock market. *Heliyon*, 5(8), e02310.
- Tesfamariam, S., & Liu, Z. (2013). *Handbook of Seismic Risk Analysis and Management of Civil Infrastructure Systems*. Woodhead Publishing Limited.
- Thmaz, O., & C.Medin, S. (1 de junio de 2019). *Passive Investment & Algorithmic Trading*. Obtenido de Research Gate: [https://www.researchgate.net/publication/333533843\\_Passive\\_Investment\\_Algorithmic\\_Trading](https://www.researchgate.net/publication/333533843_Passive_Investment_Algorithmic_Trading)
- Universidad de Pensilvania. (2020). R-squared and Root Mean Squared Error (RMSE). Obtenido de *Fundamentals of Quantitative Modeling*: <https://www.coursera.org/lecture/wharton-quantitative-modeling/4-4-r-squared-and-root-mean-squared-error-rmse-Nndhc>
- Verikas, A., Vaiciukynas, E., Gelzinis, A., Parker, J., & Olsson, M. (2016). Electromyographic Patterns during Golf Swing: Activation Sequence Profiling and Prediction of Shot Effectiveness. *Sensors*, 16(592).
- Wanjawa, B. W., & Muchemi, L. (2014). ANN Model to Predict Stock Prices at Stock Exchange Markets.
- Yu, L., Wang, S., & Keung Lai, K. (2005). Mining Stock Market Tendency Using GA-Based Support Vector Machines. *Internet and Network Economics* (págs. 336-345). Hong Kong: Springer.

## 8. ANEXO

### ANEXO I

Fórmulas medidas de precision (Yu, Wang, & Keung Lai, 2005):

$$Accuracy \text{ (exactitud)} = \frac{tp + tn}{tp + tn + fp + fn}$$

$$Precision \text{ (precisión)} = \frac{tp}{tp + fn}$$

$$Recall \text{ (sensibilidad)} = \frac{tp}{tp + fp}$$

$$Specificity \text{ (especificidad)} = \frac{tn}{tn + fp}$$

TP (*true positive*). Positivos correctamente clasificados como positivos.

TN (*true negative*). Negativos correctamente clasificados como negativos.

FP (*false positive*). Negativos incorrectamente clasificados como positivos.

FN (*false negative*). Positivos incorrectamente clasificados como negativos.

## ANEXO II

Fórmulas Indicadores Técnicos Tabla 8 (Qiu & Song, 2016):

$$OBV_T = OBV_{t-1} + \theta * V_t$$

$$MA_5 = (\sum_{i=1}^5 C_{T-i+1})/5$$

$$BIAS_6 = \left( \frac{C_t - MA_6}{MA_6} \right) * 100\%$$

$$PSY_{12} = \left( \frac{A}{12} \right) * 100\%$$

$$ASY_5 = (\sum_{i=1}^5 SY_{t-i+1})/5$$

$$ASY_4 = (\sum_{i=1}^4 SY_{t-i+1})/4$$

$$ASY_3 = (\sum_{i=1}^3 SY_{t-i+1})/3$$

$$ASY_2 = (\sum_{i=1}^2 SY_{t-i+1})/2$$

$$ASY_1 = SY_{t-1}$$

SY representa el beneficio sobre el índice Nikkei en un momento t

ASY es la media del beneficio en los últimos n días

Vt es el volumen de negociación del índice Nikkei en el momento t

PSY es la ratio del número de períodos de subida en un período de n días

A es el número de días de subida del índice en los últimos n días