

Universidad Miguel Hernández de Elche
Facultad de Ciencias Sociales y Jurídicas de Elche
Estadística Empresarial



**Métodos de Inteligencia Artificial aplicados al
comportamiento de acciones en bolsa**

Trabajo Fin de Grado
2020-2021

Autor: Javier Caro García
Tutor: José Luis Sainz-Pardo Auñón

Índice

1.- Resumen	4
2.- Introducción	4
3.- Machine Learning y el Mercado de Valores	5
3.1.- Machine Learning	5
3.1.1.- Aprendizaje supervisado	6
3.1.2.- Aprendizaje no supervisado	8
3.1.3.- Aprendizaje por refuerzo	9
3.2.- Usos Machine Learning	9
3.3.- Métodos Machine Learning	11
3.3.1.- Método K-NN	11
3.3.2.- Método SVM	13
3.3.3.- Método RL	15
3.4.- Ibex 35	16
3.5.- Trading	19
3.6.- Inteligencia artificial aplicado al mercado de valores	22
4.- Objetivos	25
5.- Metodología	28
5.1.- Herramientas	28
5.2.- Datos utilizados	29
5.2.1.- Precio de las acciones	29
5.2.2.- Obtención de los datos	30
5.3.- Funcionamiento	33
6.- Resultados	34
7.- Conclusiones	37
8.- Trabajo futuro	38
9.- Bibliografía	39



1.- Resumen

La temática del presente proyecto se engloba en predecir el comportamiento del mercado de valores o lo que se conoce comúnmente como bolsa. Para predecir la tendencia del mercado usaremos diversos métodos de Machine Learning como K-Nearest Neighbour, Support Vector Machine y Regresión Logística. Dichos algoritmos son métodos de clasificación supervisada donde analizando el comportamiento previo de una acción intentaremos predecir si la cotización de la acción subirá o bajará para poder invertir en ella de una forma más segura que con métodos de análisis técnico o fundamental. Para poder usar el algoritmo, habrá que transformar el precio en otras variables como son el rendimiento, la variabilidad y por último el volumen de operaciones para poder dar sentido al algoritmo. También se hará una breve revisión de mecanismos y funcionamientos del mercado de valores como de las últimas técnicas de inteligencia artificial usadas en los mercados financieros.

2.-Introducción

El siguiente trabajo se enmarca en el ámbito del Inteligencia Artificial, siendo este mismo muy amplio, aunque nosotros estamos únicamente interesados en el comportamiento de esta con respecto a los datos y su capacidad de aprendizaje. Esta capacidad de aprendizaje y dichos métodos que la llevan a cabo se denomina Machine Learning. El Machine Learning se nutre de los datos procesados y aprende de ellos, creando y reconociendo patrones con los cuales desarrollar soluciones sofisticadas como lo haría un humano, pero tienen una gran ventaja no necesitan descansar. Aquí es donde se puede aprovechar al máximo su potencial ya que la bolsa se mueve bajo una serie de patrones que necesitan la toma de decisiones en cuestión de segundos, lo que a un analista humano le llevaría más tiempo o incluso se le pasarían por alto ciertos indicadores estadísticos. También mediante el algoritmo sabremos cuando entrar al mercado o salir del mismo, ya que sin experiencia previa y sin los conocimientos adecuados según la CNMV el 85% de los pequeños inversores pierde parcial o totalmente el valor de sus inversiones. La Comisión Nacional del Mercado de Valores (CNMV) es un organismo regulador independiente adscrito a la secretaria de Estado de economía y Apoyo a la empresa del Ministerio de Asuntos Económicos y Transformación Digital. Fundado en 1988, se encarga de la supervisión de los mercados de valores en España

3.-Machine Learning y el mercado de valores

El Machine Learning es una técnica de Inteligencia Artificial que, a través de algoritmos, dota a los ordenadores de la capacidad de identificar patrones masivos para hacer predicciones. Dicha Inteligencia Artificial se basa en la combinación de algoritmos planteados con el propósito de crear máquinas que presenten las mismas capacidades que el ser humano. En la actualidad a esta técnica se le está sacando mucho partido. Debido a su capacidad de aprender, se puede aplicar en diversos ámbitos, pero a nosotros únicamente nos interesa su interacción con el mercado de valores y como se sirve de su aprendizaje automático para basar sus predicciones y obtener los mejores resultados.

3.1-Machine Learning

La idea de usar inteligencia artificial en este trabajo viene del creciente uso de esta tecnología, este auge viene sobre todo precedido por el aumento de la computación de los ordenadores, ya que hace relativamente poco tiempo los ordenadores no tenían la potencia necesaria para poder procesar este tipo de algoritmos ni la gran cantidad de información que necesitan para ser operativos. Hoy en día, se pueden programar para que puedan procesar gran cantidad de datos de forma autónoma. También ha cambiado el modo de programación de los algoritmos ya que tradicionalmente se usaban como inputs los datos y el modelo en cuestión para obtener los resultados. Como se puede evidenciar en la figura 1 donde primero se introducen los datos junto a un modelo en la computadora con la intención de obtener un resultado.

Figura 1. -Paradigma programación tradicional



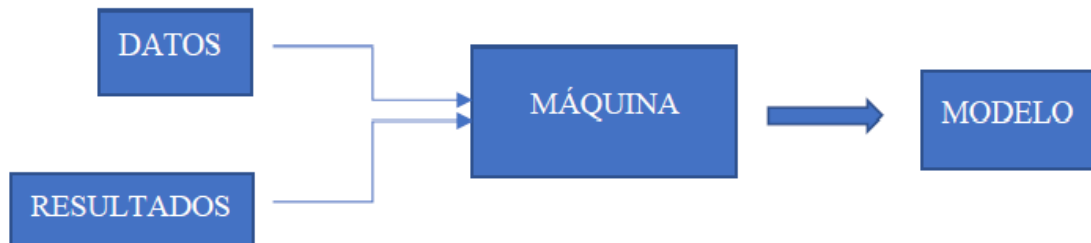
Fuente: MathWorks

Pero con el uso de la inteligencia artificial, en concreto de Machine Learning, es crear un algoritmo que revisa los datos y es capaz de predecir comportamientos futuros, basándose en los ya experimentados. El cambio de procedimiento se puede evidenciar

Métodos de Inteligencia Artificial aplicados al comportamiento de acciones en bolsa

en la figura 2, donde como inputs entran los datos y los resultados esperados, para poder obtener un output que es un modelo para usar con nuevos datos.

Figura 2.- Paradigma de Machine Learning



Fuente: MathWorks

Como hemos comentado anteriormente el Machine Learning se basa en el aprendizaje sobre los datos pasos para predecir los futuros. Dependiendo de qué tipo de métodos queramos usar existen diferentes tipos de aprendizaje el supervisado, en el que ya se tiene conocimiento de los resultados que se van a obtener y el no supervisado, que no se conocen los resultados.

3.1.1 Aprendizaje supervisado

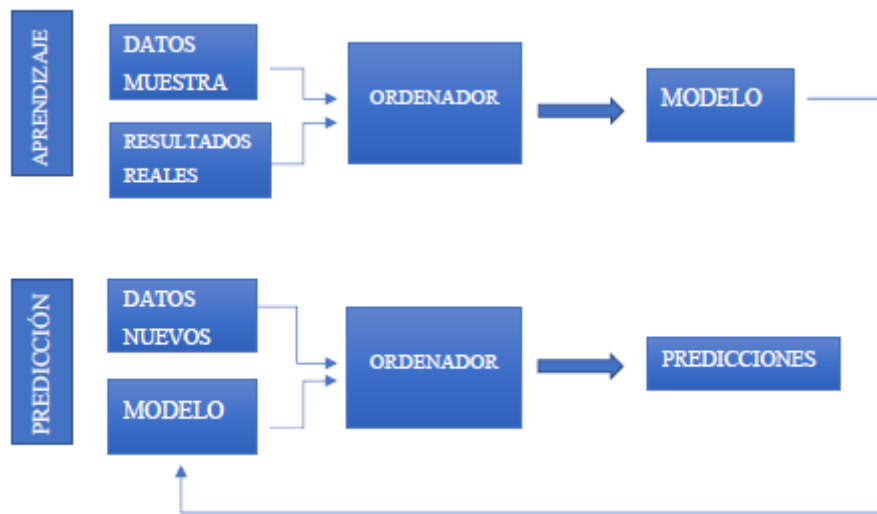
El aprendizaje supervisado genera un modelo a partir de los datos que ya conoce, es decir se le pasa una serie de datos con el resultado de la predicción para que así pueda predecir el resultado de los datos nuevos.

Dentro del aprendizaje supervisado encontramos dos técnicas:

- **Clasificación:** se utiliza con datos con respuestas categóricas, datos que tienen respuestas que ya se conocen como por ejemplo si sube o baja la bolsa.
- **Regresión:** se usa con valores continuos, en este caso contamos con una serie de variables explicativas y una variable continua, donde se intenta dar explicación a la variable continua a partir de las explicativas.

Este tipo de aprendizaje utiliza dos fases la fase de aprendizaje y la de predicción sin importar la categoría. Como evidencia la figura 3:

Figura 3.- Proceso del aprendizaje supervisado



Fuente: MathWorks

El primer paso del proceso es que el algoritmo recabe información sobre el patrón que siguen los datos, se usan los datos de entrenamiento para que cree un modelo que explique los resultados. Y se van realizando cambios hasta que el modelo sea lo más preciso posible. Una vez se ha conseguido un modelo válido se predicen los datos nuevos.

Dependiendo de cuál sea el resultado que intentemos conseguir, podemos usar los siguientes métodos según se puede ver en la tabla 1 y la tabla 2:

- **Tabla 1.- Métodos Clasificación**

Método	Descripción
Support Vector Machine	Se verá a continuación.
Clasificación Bayesiana	Dicha clasificación asume que todos los atributos de un objeto son estadísticamente independientes a una clase.
K-Nearest Neighbour	Se verá a continuación
Análisis discriminante	En este tipo de clasificador se intenta buscar relaciones lineales entre las variables continuas

	que mejor discriminen en los grupos dados a los objetos.
--	--

Fuente: elaboración propia

• **Tabla 2.- Métodos de Regresión**

Métodos	Descripción
Regresión Lineal	Es una técnica de modelado estadístico que se emplea describir una variable respuesta en función de una o varias variables predictoras.
Árboles de decisión	Según un conjunto de datos se fabrican diagramas de construcciones lógicas, sirven para representar y categorizar una serie de condiciones que ocurren de forma sucesiva.
Ensemble methods	Esté método es un conjunto de diferentes modelos y predicciones que se combinan para obtener el mejor resultado.
Redes Neuronales	Son modelos inspirados en el funcionamiento del cerebro humano. Está formado por un conjunto de nodos conocidos como neuronas artificiales que están conectadas hasta generar una salida.

Fuente: elaboración propia

3.1.2 Aprendizaje no supervisado

En el aprendizaje no supervisado los datos no están clasificados, es decir, no se tiene los resultados del modelo. A pesar de no tener información adicional el algoritmo busca patrones, estructuras y relaciones dentro de los datos con la finalidad de conseguir crear grupos que permitan clasificar un nuevo dato en función de la clasificación de los anteriores.

Entre las técnicas de aprendizaje no supervisado distinguimos:

- **Clusterización:** es una técnica exploratoria que consiste en agrupar los inputs según sus características, sin necesidad de tener conocimiento de su

estructura. Cada grupo es homogéneo entre sí y diferente con los demás grupos.

- **Reducción de dimensiones:** los datos vienen con una alta dimensionalidad, es decir, muchas variables. Con este método intentamos reducir el número de dimensiones sin perder explicación de estas, a partir de las dimensiones principales.
- **Reglas de asociación:** busca encontrar patrones que expliquen la repetición de características que se presentan simultáneamente. De esta forma se puede afirmar con cierto nivel de confianza que si una característica aparece la otra también lo hará.

Según el objetivo de nuestro análisis utilizaremos la variedad de técnicas con el fin de obtener los resultados esperados.

3.1.3 Aprendizaje por refuerzo

El aprendizaje por refuerzo se basa en recompensar o castigar al algoritmo, es decir, otorgarle al algoritmo un refuerzo positivo cada vez que aciertan o penalizarlo cada vez que falla. Con este tipo de acciones lo que se intenta es que el algoritmo aprenda de sus fallos y de sus aciertos para ir perfeccionando el algoritmo y sea lo más eficiente.

3.2. Usos del Machine Learning

Las herramientas de Inteligencia Artificial como en su caso el Machine Learning permite a las entidades analizar un volumen de datos mucho mayor que con otras herramientas tradicionales o de una tecnología inferior y de manera más rápida. Además, el aumento de variables a procesar lleva a un análisis de mayor calidad dado que consigue conocer mejor la temática investigada y como consecuencia resultados más precisos.

Estos son algunos de los ejemplos de aplicaciones técnicas de la Inteligencia Artificial, véase [5], que están creciendo en la actualidad:

- **Reconocimiento de imágenes estáticas, clasificación y etiquetado.** Estas herramientas son útiles para una gran gama de industrias.

Métodos de Inteligencia Artificial aplicados al comportamiento de acciones en bolsa

- **Mejoras del desempeño de la estrategia algorítmica comercial.** Ya ha sido implementada de diversas maneras en el sector financiero donde se usa para la clasificación de clientela.
- **Detección y clasificación de objetos.** Se usa en la industria de vehículos, donde se busca la conducción autónoma del vehículo.
- **Distribución de contenido en redes sociales.** Se trata en su mayoría de técnicas de marketing utilizadas en las redes sociales, aunque podría usarse en distintos ámbitos.

Ahondando más en el sector financiero, la aplicación de técnicas de inteligencia artificial en la prestación de servicios puede dar lugar a mejoras de eficiencia, reducción de costes, incrementos de calidad, aumento de nivel de satisfacción de los clientes, gracias, principalmente a la posibilidad de automatizar procesos operativos e incrementar capacidades analíticas. No todo en estas aplicaciones es bueno también hay que llevar cuidado con el sesgo presente en los algoritmos y con la forma de razonar de estos, a veces se hace muy complicado conocer como han llegado a dicha conclusión. Siendo estos algoritmos en ocasiones cajas negras, es decir, no sabemos cómo llegan exactamente a esa conclusión o su pensamiento.

El otro ámbito de nuestro estudio es el mercado de valores o lo que comúnmente se le denomina “bolsa”. La bolsa es un conjunto de instituciones y agentes financieros que negocian sobre los distintos tipos de activos, como lo son:

- **Las acciones** son títulos emitidos por una Sociedad Anónima o Sociedad comanditaria por acciones que representa el valor de una de las fracciones iguales en la que se divide su capital social.
- **Las obligaciones** son emisiones de deuda que realiza una empresa o una entidad pública. Mediante la compra, presta dinero al emisor a cambio de la devolución principal más un tipo de interés.
- **Los fondos** son instituciones de Inversión colectiva, es decir, es un instrumento de ahorro que aglutina las aportaciones hechas por un número de personas que quieren invertir sus ahorros.

Se mueven gran cantidad de acciones día a día en todo el mundo, con la finalidad de conseguir una plusvalía por la venta de los activos. En este sector existe la creencia por partes de la sociedad que no conocen el funcionamiento de este sistema que es como

“jugar” a la lotería y que no tiene ningún fundamento. A pesar de que no hay ningún método que haya sido efectivo en su totalidad debido a su gran componente de aleatoriedad, aunque como dice Andrew W. profesor del Instituto Tecnológico de Massachusetts [22], sí que se ha demostrado que el mercado de valores sigue patrones que se pueden evidenciar bajo el análisis técnico, fundamental o de algún otro algoritmo basado en mecanismos estadísticos que ha dado sus frutos. Por este motivo vamos a intentar predecir su comportamiento con los diferentes métodos de Machine Learning que hemos mencionado anteriormente que son K-Nearest Neighbour, Support Vector Machine y Regresión Logística.

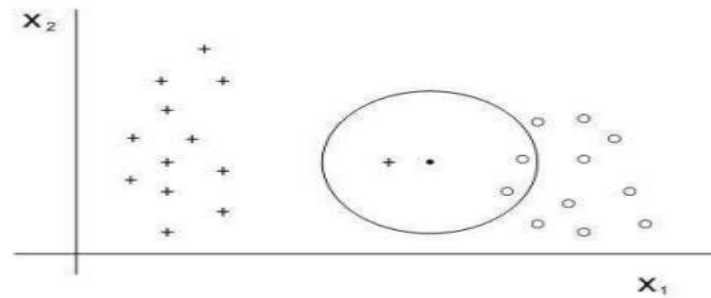
3.3- Métodos de Machine Learning

En los siguientes apartados vamos a comentar y explicar que métodos de Machine Learning en este caso métodos supervisados, hemos usado para llevar a cabo nuestro algoritmo.

3.3.1.- Método K-NN

Por la parte del método estadístico tenemos el método de K-Nearest Neighbour (K-NN), como nos indica el profesor José Luis Sainz [17], este método se englobaría dentro del Machine Learning, una rama de la Inteligencia Artificial. Es un método de clasificación supervisada. La idea es bastante simple dado un individuo a clasificar, se localiza en los datos de muestra o entrenamiento (cuya clasificación se conoce) los k individuos más cercanos al individuo a clasificar. Finalmente, el individuo se clasifica en la categoría a la que pertenece la mayoría de los k individuos más cercanos. Un ejemplo aportado por Abdelmalik Moujahid, Iñaki Inza y Pedro Larrañaga [20] de cómo funciona este algoritmo se puede ver en la figura 4. Donde tenemos 24 casos ya clasificados en dos posibles valores ($m=2$). Las variables predictoras son X_1 y X_2 , y se han seleccionado $K = 3$. De los 3 casos ya clasificados que se encuentran más cerca del nuevo caso a representar x representado por un punto en el centro de la imagen. Dos de los individuos más cercanos son de la clase círculo por ello el algoritmo con $K=3$ lo clasifica con la clase círculo. Por el contrario, si hubiera sido $K=1$ se hubiera quedado con la cruz ya que es el vecino más cercano.

Figura 4.- Ejemplificación de uso del K-NN.



Fuente: Abdelmalik Moujahid, Iñaki Inza y Pedro Larrañaga [20].

Existen una serie de variantes del algoritmo K-NN. Las variantes se usan para poder corregir ciertos fallos que pueden presentar el método en ciertas particularidades estas son las más representativas:

- **K-NN con rechazo**, la idea de esta variante es bastante simple que si el algoritmo no tiene garantías de como clasificar un individuo lo rechaza y lo deja sin clasificar.
- **K-NN con distancia media**, este método se caracteriza por clasificar el individuo a la clase que menor distancia tenga de él.
- **K-NN con distancia mínima**, dicha variación consiste en seleccionar un caso por clase, normalmente el caso más cercano al baricentro de todos los elementos de dicha clase. En este paso se reduce la dimensión del fichero de casos a almacenar de N a m. Por último, se asigna el nuevo caso a la clase cuyo representante está más cercano. El funcionamiento del método depende en mayor parte de la homogeneidad de sus clases.
- **K-NN con pesado de las variables**, dar el mismo peso a todas las variables puede resultar poco eficiente para el método en el caso de que algunas de las variables sean irrelevantes para cierto individuo. Por este motivo resulta eficiente usar distancias entre casos que ponderan cada variable de manera adecuada. La forma de calcular la distancia sería la siguiente:

$$d(\mathbf{x}, \mathbf{x}_r) = \sum_{j=1}^n w_j (x_j, x_{rj})^2$$

Con lo cual la variable X_j tiene asociado un peso w_j que habrá que determinar.

El método K-NN que vamos a utilizar presenta una serie de observaciones:

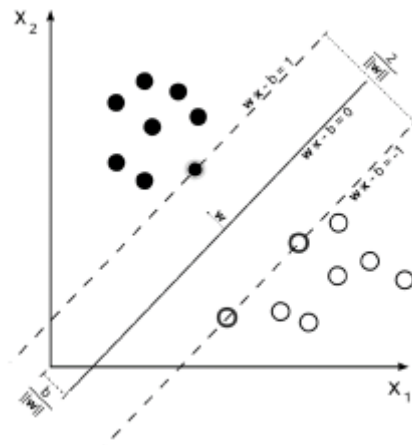
- Dado que tiene que haber una mayoría se suele escoger k con valores impares: 1,3,5....
- ¿Qué valor de K coger? Se suele probar con distintos k y el que mayor precisión aporte con los datos de entrenamiento.
- Se pueden usar distintas distancias: Euclídea, Manhattan, Máxima, Mínima, Mahalanobis.... Al igual que con las k se prueba con varias.
- En caso de tener una gran cantidad de individuos a clasificar, es decir, miles o millones se suele escoger los individuos más significativos, a estos individuos se les llama prototipos. Es una forma de poder clasificar los individuos con mayor rapidez.
- Un método para seleccionar estos individuos puede ser el método de las K -medias (clúster) formando tantos grupos como individuos se desea seleccionar y usar los centroides de estos grupos como prototipos.

3.3.2.- Support Vector Machine

Las máquinas de vector soporte o como se conoce por su término en inglés Support Vector Machines (SVM) según Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani [6] y también el profesor José Luis Sainz [18], es un tipo de algoritmo de clasificación supervisada. El SVM es una generalización de un clasificador simple llamado maximal margin classifier. Sin embargo, el SVM permite una mejor clasificación para casos con límites no lineales entre clases.

El funcionamiento del SVM es bastante sencillo y consiste en formular un hiperplano que separe de forma óptima a los puntos de una clase de la otra, de forma que exista la máxima distancia o margen entre los puntos más cercanos al mismo. En la figura 5 se puede evidenciar un claro ejemplo de este método.

Figura 5.- Funcionamiento SVM



Fuente: apuntes José Luis Sainz-Pardo Auñón

El concepto del hiperplano es fundamental en este método ya que es su base. Un hiperplano es un subespacio plano y afín de dimensión $p-1$, que divide el espacio en dos mitades. Por ejemplo, en un espacio de 2 dimensiones un hiperplano es un subespacio plano de una sola dimensión o lo que es lo mismo una línea definida por la ecuación:

$$\beta_0 + \beta_1 x_1 + \beta_2 x_2 = 0$$

Donde cualquier $X = X_1, X_2$ para los que se cumple la ecuación es un punto en el hiperplano. El ejemplo anterior también se puede extender a 3 dimensiones.

Clasificación para casos no separables literalmente, es una posibilidad para tratar con límites no lineales entre clases consiste en aumentar manualmente el espacio de los predictores mediante funciones polinómicas o con términos de interacción. Para evitar el uso de demasiados predictores usamos el Vector Support Machine que es una extensión como hemos comentado anteriormente de los support vector classifiers que aumenta la dimensionalidad a partir del uso de kernels, un enfoque computacionalmente más eficiente. Los kernels son funciones que transforman un espacio de pocas dimensiones en un espacio de dimensiones mayores mediante transformaciones complejas de los datos.

Los kernels más utilizados son:

- **Kernel lineal**, este kernel cuantifica la similitud en un par de observaciones usando la correlación de Pearson.

- **Kernel polinómico**, permite un límite de decisión mucho más flexible. Cuando un Support Vector Classifier se combina con un kernel no lineal se consigue el Support Vector Machine.
- **Kernel radial**, tiene un comportamiento muy local, en el sentido de que solo las observaciones de entrenamiento cercanas a una observación de test tendrán efecto en su clasificación.

Cuando hablamos de flexibilidad no hablamos de una mejor predicción ya que puede haber un sobreajuste en los datos de entrenamiento. Dicho sobre ajuste se produce cuando un modelo se entrena con demasiados datos, el algoritmo empieza a aprender del ruido y de entradas inexactas en nuestros conjuntos de datos lo que provoca fallos en la predicción.

En el caso de querer clasificar en $K > 2$ es decir en más de dos categorías, necesitaremos obtener $k-1$ Hiperplanos soporte, existiendo diversos métodos. Uno de dichos métodos, el llamado uno contra el resto consistirá en utilizar $k-1$ clasificadores con valores 1 si el individuo pertenece a un grupo determinado, y 0 si pertenece al resto. En caso de que el discriminador asigne un individuo a dos o más grupos, se decidirá el valor absoluto del discriminador más alto.

3.3.3.- Regresión Logística

La regresión logística, como plasma José Luis Sainz en su trabajo como docente e investigador [19], es un tipo de análisis de regresión utilizado para predecir el resultado de una variable categórica. Por tanto, puede utilizarse también como un método de clasificación supervisada.

Se enmarca en el llamado conjunto de Modelos Lineales Generalizados y se puede formular como:

$$y = f(z) = f(w_0x_0 + \dots + w_nx_n)$$

Podríamos utilizar distintas funciones $f(z)$ como funciones discriminantes. Una de ellas, la función logística o sigmoide que viene expresada por:

$$f(z) = \frac{1}{1+e^{-z}}$$

La función logística posee las siguientes propiedades que la convierten en muy útil como función discriminante:

- Está acotada entre 0 y 1.
- Podemos, por tanto, interpretar sus resultados como probabilidades.
- Para problemas de clasificación binaria como el nuestro podemos suponer que los valores menores de 0.5 corresponden a la clase 0 y los superiores a la clase 1.

3.4.- IBEX 35

Aunque con nuestro algoritmo podamos usar como datos predictivos cualquier valor que presente cotización previa al estudio. Centraremos nuestros esfuerzos en comentar el índice bursátil más relevante en nuestro país, el Ibex 35. En España hay 4 bolsas físicas, la más importante se encuentra en Madrid y las otras 3 en Barcelona, Valencia y Bilbao. Un índice bursátil es un sistema de medida estadístico con el cual poder ver la evolución de las acciones cotizadas en dicho índice a lo largo del tiempo. El Ibex 35 se fundó en 1992 con 3000 puntos como base, normalmente se compone de 35 empresas de ahí su nombre, aunque ha llegado a ser menor este número como también mayor como se puede ver en la figura 6 su evolución desde su fundación.

Figura 6: evolución histórica del Ibex 35



Fuente: elaboración propia

Métodos de Inteligencia Artificial aplicados al comportamiento de acciones en bolsa

En él cotizan las empresas con más liquidez presentes en el Sistema de Interconexión Bursátil Español. La empresa encargada de la gestión del mercado español es Bolsas y Mercados Españoles o su acrónimo BME [16]. Un comité de expertos se reúne al menos 2 veces al año para revisar la composición. También dentro del Ibex 35 hay una variedad de subíndices como pueden ser el Ibex35 con dividendos, Ibex medium CAP, Ibex 35 Bancos etc. La finalidad de estos subíndices es la exploración de nuevos productos financieros fijados en ciertos valores del índice. Se pueden ver en la tabla 3 las acciones cotizadas en el índice actualmente, su sector y su capitalización.

Tabla 3.- Componentes del Ibex 35 y sus ponderaciones

Nombre	Ticker	Sector	Ponderación
IAG	IAG.MC	Transporte	2.68
Sabadell	SAB.MC	Finanzas	0.59
Telefónica	TEF.MC	Telecomunicaciones	4.98
Repsol	REP.MC	Energía	3.37
Merlin Properties	MRL.MC	Inmuebles	0.93
Meliá Hotels	MEL.MC	turismo	0.27
Santander	SAN.MC	Finanzas	11.19
BBVA	BBVA.MC	Finanzas	6.96
Bankinter	BKT.MC	Finanzas	1.12
Indra	IDR.MC	Consultoría	0.30
Mapfre	MAP.MC	finanzas	0.74
Colonial	COL.MC	Inmuebles	0.56
CaixaBank	CABK.MC	Finanzas	3.43
ACS	ACS.MC	Infraestructuras	1.88
Almirall	ALM.MC	Farmacéutica	0.39
Aceronix	ACX.MC	Metalurgia	0.63
Amadeus IT Group	AMS.MC	Turismo	5.99
Naturgy	NTGY.MC	Energía	1.79
Grifols	GRF.MC	Farmacéutica	2.08
ArcelorMittal	MTS.MC	Metalurgia	1.14
Aena	AENA.MC	Transportes	3.78
Inditex	ITX.MC	Textil	11.63

Métodos de Inteligencia Artificial aplicados al comportamiento de acciones en bolsa

Enagás	ENG.MC	Energía	1.05
Ferrovial	FER.MC	Infraestructuras	3.44
CIE Automotive	CIE.MC	Automoción	0.5
Red Eléctrica Española	REE.MC	Energía	1.76
Endesa	ELE.MC	Energía	2.09
Acciona	ANA.MC	Infraestructuras	0.99
Viscofan	VIS.MC	Alimentación	0.58
Iberdrola	IBE.MC	Energía	15.61
Cellnex Telecom	CLNX.MC	Telecomunicaciones	3.87
Siemens Gamesa	SGRE.MC	Tecnológica	1.69
Pharmamar	PHM.MC	Farmacéutica	0.42
Solaria	SLR.MC	Energía	0.49

Fuente: elaboración propia

La distribución de las empresas del Ibex35 quedaría de la siguiente forma:

1. Petróleo y energía → 26.52%
2. Materiales básicos, industria y construcción → 10.26%
3. Bienes de consumo → 15.11%
4. Servicios de consumo → 6.73%
5. Servicios financieros → 24.75%
6. Tecnología y telecomunicaciones → 15,14%
7. Servicios inmobiliarios → 1.49%

La fórmula para el cálculo del índice es:

$$I(t) = I(t - 1) \times \frac{\sum_{i=1}^{35} Cap_i(t)}{[\sum_{i=1}^{35} Cap_i(t - 1) \pm J]}$$

donde I(t) es el valor del índice en el momento t, I(fecha de fundación)=2676.12, Cap es la capitación bursátil del *free float*, cuyo significado hace referencia a las acciones que están en circulación por parte de las compañías que integran el índice y J es un coeficiente usado para ajustar el índice para que no se vea afectado por ampliaciones de capital.

3.5.- Trading

El trading se puede definir como las transacciones a corto o medio plazo con el objetivo de ganar una plusvalía. Los traders suelen aplicar diferentes métodos y técnicas en su operativa bursátil, entre ellos el análisis técnico. Este tipo de análisis se basa en el estudio de la acción del mercado, principalmente a través del uso de gráficos con el propósito de predecir la futura evolución del precio del activo. Con el estudio de los gráficos se intenta adelantar a los movimientos del precio y así entrar en el momento correcto al mercado. El precio en este tipo de técnica es la variable más importante.

Hay que especificar que los teóricos en la materia dividen el análisis técnico en dos grandes grupos como indica Francisca Serrano en su libro[7]:

- **Análisis de gráficos chartista**, un chartista es la persona que se dedica profesionalmente a analizar y predecir los valores de las acciones y sus variaciones en bolsa a través de los charts ,es decir, los gráficos. En este caso, se buscan las figuras geométricas que el precio va creando en el gráfico con sus movimientos alcistas, bajistas o laterales. No se usan herramientas adicionales.
- **Análisis técnico en sentido estricto**, el estudio se centra en la observación de los indicadores, retrasados o adelantados, nos permitirán prever que hará el precio en el futuro.

A parte de usar el machine learning para la inversión, también se puede usar en otros ámbitos como el riesgo de los activos o las fronteras eficientes de inversión.

Para poder usar el trading es necesario tener conocimientos de mercados financieros, los activos de dichos mercados y los riesgos de la operativa bursátil. Muchos de los expertos en trading coinciden en que esta práctica reduce el riesgo en la inversión ya que nunca se deja la posición abierta mucho tiempo y no suelen ir contra el mercado, sino que lo acompañan. Estos son los tipos de traders más comunes:

- **Scalping o quick trading**. Los scalpers ejecutan una operativa que se lleva a cabo en minutos o segundos se suelen utilizar gráficos de uno, tres o cinco minutos. Esta técnica tiene muchos seguidores por la alta volatilidad del mercado. Su objetivo es localizar valores que durante la sesión sufran fuertes caídas o subidas en un corto periodo de tiempo.

Métodos de Inteligencia Artificial aplicados al comportamiento de acciones en bolsa

- **Day trading**, este tipo de trader se lleva en un marco superior a la hora hasta antes del cierre del mercado. Usa gráficos de 15,30, 60 o 240 minutos. El objetivo es predecir una fuerte subida o caída, pero a diferencia del quick trading acompaña todo el movimiento de la gráfica.
- **Traders de posición**, aguantan las posiciones al cierre del mercado. La posición suele quedar abierta porque está en beneficio y bien protegida.
- **Swing trading**. Aquí nos olvidamos de las técnicas intradiarias e intentamos cazar la tendencia principal del gráfico dejando de lado las fluctuaciones de precio a corto plazo.
- **Trading de alta frecuencia**, este tipo de trading no lo llevan a cabo personas, sino que aprovecha la potencia de los procesadores informáticos para lanzar ordenes automatizadas que se suelen adelantar al pequeño inversor. La negociación de alta frecuencia o en inglés *high-frequency trading* emplea algoritmos informáticos con los que analiza el mercado, entonces los ordenadores plantean una estrategia de inversión automatizada.

Una vez conocidos los tipos de traders, es relevante conocer ciertos indicadores que usan los traders como herramientas de ayuda a la inversión, estos indicadores son fórmulas matemáticas y estadísticas que se aplican a las series de precios y volúmenes con la intención de ayudar a tomar decisiones de inversión o a ubicar a los precios de determinadas fases o situaciones. Se distinguen 2 tipos:

- **Indicadores retrasados**: aquellos que son más precisos durante los períodos en los que existe tendencia y están basados en medias móviles.
- **Indicadores adelantados**: aquellos que ofrecen señales más fiables durante los períodos en que el mercado se mueve en rasgos, en mercados laterales sin una tendencia definida.

También es importante conocer los tipos de mercados existen en el mundo bursátil y como se opera en cada uno de ellos. Se pueden diferenciar tres tipos de mercados:

- **Mercado primario**: es el mercado que por primera vez emite un activo financiero. Se suele invertir en este mercado mediante ofertas preferentes de compra, es decir, tienes una opción de poder comprar dicha acción antes que los demás. Por otra parte, también se suelen realizar ofertas públicas de adquisición donde cualquier inversor puede adquirir la acción recién emitida.

Métodos de Inteligencia Artificial aplicados al comportamiento de acciones en bolsa

- **Mercado secundario:** donde se realiza la transacción de valores que previamente han sido emitidos en el mercado primario, el ejemplo principal es el mercado de valores.
- **OTC:** Viene de las siglas de Over the Counter, se conoce así a los mercados no organizados. Estas operaciones se llevan a cabo directamente entre empresas interesadas en recibir financiación.

Por otra parte, dentro de los mercados financieros tenemos los activos financieros que son los objetos con los que negociamos en el mercado para obtener una plusvalía. Estos activos tienen diferentes usos y mecanismos de funcionamiento. Los principales son:

- **Renta fija y variable:** acciones y bonos, private equity.
- **Commodities:** materias primas, sector inmobiliario, petróleo, minerales.
- **Divisas:** moneda extranjera y tipos de cambio, criptomonedas.

En los siguientes activos se puede invertir directamente a través de un bróker o a través de terceras personas como fondos de inversión, fondos de inversión cotizados o mediante sociedades de capital variable o derivados.

Otro de los elementos a tener en cuenta en la inversión es el riesgo tomado con cada activo o inversión realizada. El riesgo hace referencia a la posible incertidumbre de conseguir el resultado deseado. El riesgo en los mercados lo suele medir las agencias calificadoras de riesgo indicando mediante su criterio cada nivel de riesgo de crédito, es decir, el impago de ese activo. En España a partir de 2016 entró en vigor una nueva fórmula para determinar el riesgo de un activo financiero. De esta nueva forma se ha intentado simplificar la nomenclatura para hacerla más accesible a todos los públicos. El sistema clasifica los activos financieros según su riesgo del nivel 1, el más bajo hasta el 6, el más alto como se muestra en la figura 7.

Figura 7.- Riesgo de activos financiero



Fuente: Boletín Oficial del Estado

3.6. - Inteligencia artificial aplicada al mercado de valores

Aparte de conocer cómo funcionan los algoritmos K-Nearest Neighbour, el Support Vector Machine y la Regresión Logística también nos es de interés aprender más sobre como este tipo de tecnologías se aplica a los mercados financieros y cómo esta ayuda a mejorar el rendimiento de las inversiones. Estos sistemas de algoritmos son relativamente recientes, puesto que lo que se usaba anteriormente son los sistemas de trading automático clásicos que basan sus reglas en la estrategia que el trader ha desarrollado a lo largo del tiempo con su experiencia negociando en los mercados. No son malos sistemas siempre y cuando se compruebe estadísticamente que no son solo suerte y realmente funciona.

Los sistemas basados en modelos [9] como el nuestro son diferentes. Lo que hacen los modelos son analizar la información disponible buscando un patrón que con alta probabilidad continúe en determinado momento del mercado. Para ello se han utilizado diversas tecnologías, pero casi todas tienen algo en común la Inteligencia Artificial. El creciente uso de las redes a la hora de realizar las transacciones está generando un cambio de tendencia en el uso de nuevas tecnologías a la hora de gestionar activos y vayan tomando la decisión los algoritmos de computación ya que en segundos o pocos minutos pueden ser capaces de generar una estrategia de inversión adecuada.

Para que estos algoritmos puedan funcionar con las mayores garantías deberían de usar información a tiempo real mediante técnicas de Big Data que según De Mauro, Greco y Grimaldi (2014)[10] lo definen como:

“ El Big Data representa los activos de información caracterizados por un alto volumen, velocidad y variedad que requieren de una tecnología específica y de métodos analíticos para transformarla en valor”

Aplicar técnicas de Big data en los mercados bursátiles, como se puede evidenciar en el blog BIIC[11], permitiría mejorar los procesos si se consigue el reto de analizar de forma eficaz la ingente cantidad de datos que generan los mercados financieros disponibles y la información que de ellos se desprende tras haber limpiado el ruido de los datos de poca calidad. El estudio de toda esta gran cantidad de datos puede aportar información clara y concisa no visible a primera vista.

Métodos de Inteligencia Artificial aplicados al comportamiento de acciones en bolsa

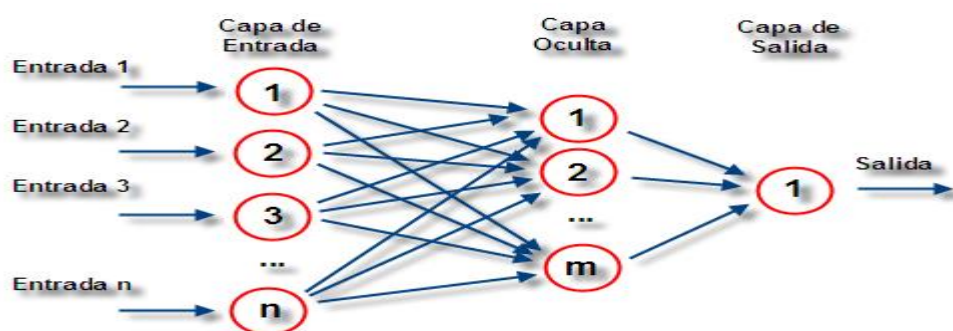
Una de las ventajas del Big Data, como dicen en su libro Sudeep Tanwar, Sudhansu Tyagi Neeraj Kumar [15], es que puede tratar toda la información que se genera a una gran velocidad. Saber cómo reaccionar en tiempo real y de forma proactiva a las más o menos inesperadas subidas y caídas de los productos bursátiles es fundamental para evitar graves pérdidas y obtener beneficios.

A parte de las técnicas de Big Data que nutren de información a otros sistemas como son las redes neuronales, estas redes son un sistema de aprendizaje automático basado en la forma de funcionar del sistema nervioso de ahí su nombre. Se trata según Esteban Pérez[4], de un sistema de conexión de neuronas que colaboran entre sí para producir un estímulo de salida. En una red neuronal, recibimos los datos de entrada a través de una capa (nodos de entrada), serán procesados en otra capa (nodos ocultos). Estos nodos ocultos son fundamentales pues es donde la red neuronal aprende y realiza la predicción óptima para el problema planteado. Cada neurona esta interconectada a través de un enlace. El valor de salida de una neurona es multiplicado por un peso al llegar a la siguiente. El estado de ese peso pueda inhibir o activar las siguientes neuronas. Esta función genera que este sistema se pueda entrenar donde el sistema aprende por sí mismo. Las conexiones de pesos que conectan las redes pueden ser de 3 tipos diferentes, como indica Charu C. Aggarwal [14] en su libro :

- **Feedforward**, los resultados de las capas inferiores se dirigen hacia las capas superiores mediante conexiones hacia delante.
- **Feedback**, como su propio nombre indica los resultados de las capas superiores retroceden hacia las inferiores.
- **Lateral**, este formato asigna el valor total al valor más alto de salida, al resto de valores les asigna un cero.
- **Retardo**, es usado en modelos dinámicos y temporales, donde es necesaria memoria.

El Objetivo es que el algoritmo tome decisiones acertadas sin la intervención del ser humano. En la figura 8 se encuentra una demostración de este tipo de redes neuronales.

Figura 8.- Estructura red neuronal.



Fuente: Rankia[4]

En su funcionamiento aplicado a la bolsa no se conseguiría un 100% de efectividad debido a diversas cuestiones relacionadas con el comportamiento de los mercados no obedece a leyes exactas, es decir, su gran aleatoriedad. Su funcionamiento se asemejaría a otros modelos de Machine Learning, la red necesita una serie de datos de entrada como podrían ser: la variación de los precios de cierre de las últimas sesiones, los puntos de pivote diario, el volumen operado, etc. Además de estos datos le proporcionaremos si la sesión cerró alcista o bajista, para poder entrenarla.

Posteriormente le pasaremos una nueva serie de datos donde no le decimos si cerró al alza o a la baja, pero sí los otros datos. La red tratará de predecir cómo fue el cierre de sesión y nos dará por cada sesión un valor de confianza para la predicción alcista y otro para la bajista. Tomando el valor más alto tenemos la predicción.

Por otra parte, una de las técnicas más novedosas se basa en algoritmos naturales en este caso, como se explica en el blog slowinver [12], un algoritmo que imita el comportamiento de una bandada de estorninos, un tipo de pájaro que se mueve al unísono. Este algoritmo intenta reproducir este comportamiento de un modo simplificado, permitiendo elegir las acciones que se mueven con una correlación positiva.

En un conjunto de acciones líderes son aquellas que se mueven con más velocidad que las demás, bien al alza o a la baja. El algoritmo entra en el bloque de acciones que empuja hacia arriba la bandada y sigue entrando y saliendo sistemáticamente en las acciones que su cotización cotice al alza o a la baja si operamos a corto. El algoritmo funciona mediante ordenes de tipo rotacional, es decir, forma una cartera de acciones y se rotan cada cierto tiempo, vendiendo las que menos rendimiento ofrezcan y comprando nuevos títulos. Este algoritmo, falla cuando los valores en cartera sufren una

gran volatilidad por este motivo se le añaden filtros que limiten fuertes subidas y bajadas, haciendo que el algoritmo venda ese valor y se pase a tener el dinero líquido esperando a otra oportunidad de inversión.

Por último, no podemos olvidarnos de las predicciones que realizan las series temporales, ya que la cotización de una acción es una serie temporal. Dichas series, véase [13], se definen como una colección de observaciones de una variable recogidas secuencialmente en el tiempo. En este caso el algoritmo usado es el análisis de un solo espectro(SSA). Lo que realiza el SSA según Acelera.tech [3], es una descomposición de la serie temporal en componentes principales. Esta descomposición en componentes principales se puede interpretar como partes de la serie que corresponden a tendencias, ruido, estacionariedad y muchos otros factores, se realiza esta operación para convertir la serie en un ruido blanco. Posteriormente se reconstruye y se usan para obtener una predicción. Una peculiaridad de este método es que no se obtienen predicciones basándonos en nuevos datos como en modelos de regresión o de clasificación, sino que se obtienen con los datos pasados y realizando la predicción sobre ellos.

4.- Objetivos

El objetivo y motivación principales del proyecto es predecir si un activo en concreto en este caso acciones ya que lo vamos a implementar en el índice Ibex 35, suben su cotización en un periodo futuro, es decir, mañana, en un mes o según los parámetros con los que trabaje el algoritmo, o si de lo contrario su cotización cae. Estos métodos clasifican a los sujetos o en este caso las acciones. Según el algoritmo a usar variara la forma de encontrar la solución óptima, siendo estas, las distancias a los vecinos en el caso del K-NN, su imagen en el hiperplano en el caso de SVM o su representación mediante la función logística en el caso de la Regresión Logística, es decir, a otras acciones o cotizaciones que se comporten de una forma similar. Nos indicará con un 1 si la acción sube o con un 0 si la acción baja. Podremos invertir en ella tanto de forma alcista como abrir una posición a corto de una forma bajista. Los objetivos principales del trabajo son:

- Diseñar e implementar una herramienta que permita al usuario invertir de una forma segura, es decir, minimizando el riesgo.

Métodos de Inteligencia Artificial aplicados al comportamiento de acciones en bolsa

- Discernir entre los 3 métodos de Machine Learning que mejores resultados arrojan a nuestro algoritmo. Se utilizarán los métodos K-Nearest Neighbour, Support Vector Machine y por último Regresión Logística.
- Determinar cuáles son unas correctas variables para el algoritmo y como transformarlas para que aporten información al modelo.

Durante el proyecto trabajamos sobre la hipótesis de que las acciones se mueven entorno a patrones estadísticos que pueden ser captados y reproducido por un método automáticos. Según el BBVA [1] la formación del precio de una acción y su fluctuación dependen de ciertos factores:

- **Ley de la oferta y la demanda**, es decir, a cierto precio cuantos compradores están dispuestos a adquirir la acción. Si muchos compradores quieren esa acción esta sube de precio.
- **Los Factores fundamentales** consisten en las causas que provocan movimientos en los precios como por ejemplo noticias, decisiones macroeconómicas, eventos corporativos y resultados económicos. Los factores fundamentales más importantes son, los beneficios, la expectativa de los inversores y el mercado sobre los resultados de una empresa, ratios de descuento y por último el riesgo percibido en la acción.
- **Factores técnicos** que se basan en el estudio de datos de precios históricos para intentar predecir movimientos futuros del mercado. Esta técnica se ayuda de las gráficas que suelen generar figuras que ayudan a la predicción de la tendencia del precio y distintos indicadores como las medias móviles que se usan también con la intención de adivinar estos cambios de tendencia.
- **Los sentimientos del mercado**, se refiere a la psicología de los agentes que intervienen en el mercado. Que frente a ciertos acontecimientos de este actúan de una forma u otra.
- **Correlación**, es un nexo que mide la relación existente entre las inversiones de una cartera y el índice de referencia. Se dice que si una acción tiene una fuerte correlación con su índice o con otra acción los comportamientos de

Métodos de Inteligencia Artificial aplicados al comportamiento de acciones en bolsa

ambos se parecen. Si por el contrario tienen una correlación negativa se comportan de forma diferente, lo que permite minimizar el riesgo invirtiendo en acciones de este tipo. Ya que nunca van a bajar las dos a la vez.

La interpretación de todos estos factores nos ayudara a intentar anticipar los movimientos en el precio de los activos financieros. Pero sin olvidar que es la ley de la oferta y la demanda la que determina en mayor medida el precio de las acciones. Esta serie de factores también puede hacer fallar nuestro algoritmo ya que una noticia mala sobre una de las compañías que nuestro algoritmo diga que es propicio para invertir puede cambiar la tendencia del precio y hacer que la predicción no sea correcta. Por eso hay que tener en cuenta el resto de los factores antes de realizar la inversión.

Otro de los puntos a tener muy en cuenta es la entrada en el mercado o lo que es lo mismo comprar o vender una acción. Existen dos tipos de formas de entrar al mercado:

- **Compra de acciones**, este método suele ser el más común. Es comprar una acción a un precio dado y esperar que suba para poder sacar una plusvalía entre el precio de compra y el precio de venta.
- **Venta a corto de acciones**, este método consiste en invertir en un valor cuando se piensa que su cotización va a descender de precio y si desciende obtenemos un beneficio. Hay que llevar cuidado con este método ya que realmente no estamos comprando esas acciones, sino que las pedimos prestadas, es decir, vendemos primero para luego comprarlas más baratas y obtener así un beneficio.

Debemos de prestar también atención al intermediario financiero o bróker con el que realizamos las operaciones ya que ellos se ocupan de la transacción de las acciones entre el mercado de valores y nosotros. Por este servicio se llevan una comisión y nos convendría contratar el operador que menor comisión cobre por el servicio o el que mejor plataforma de inversión tenga a su disposición. El bróker también nos permite apalancarnos o lo que es lo mismo pedir dinero para operar, las ganancias pueden ser mucho mayores utilizando este sistema, pero las pérdidas también.

5.- Metodología

En este capítulo se aborda la descripción del proyecto a realizar para evaluar la capacidad de nuestro algoritmo para prever las subidas y bajas de las acciones cotizadas del mercado de valores y así poder invertir con mayor seguridad.

Para la realización del trabajo hemos tenido en cuenta las empresas del IBEX-35, excluyendo Bankia por su fusión con CaixaBank.

También es importante comentar, que para el algoritmo podríamos usar cualquier otro índice o empresa del mundo que haya cotizado entre el 01-01-2020 y el 10-03-2021 dado que estos son nuestros periodos de referencia. Entre dichos periodos lo que nos interesa saber cómo inversores es si la predicción de la acción es alcista o bajista, para poder decidir nuestra estrategia de inversión y como llevarla a cabo.

5.1. -Herramientas

En el desarrollo hemos usado el programa RStudio, véase [21], es un entorno de desarrollo integrado para el lenguaje de programación R, dedicado a la computación estadística y gráficos. Esta ha sido la herramienta seleccionada para llevar a cabo el algoritmo. Para ello también haremos uso de las siguientes librerías:

- **Quantmod**, es la librería usada en R para los modelos financieros cuantitativos.
- **Zoo**, librería con métodos de observaciones indexadas totalmente ordenadas. Está dirigido a series de tiempo irregulares de vectores y matrices numéricas.
- **PortfolioAnalytics**, corresponde a una librería que contiene rutinas y gráficos de optimización para el análisis carteras.
- **Xts**, corresponde a una librería de apoyo a zoo donde se proporciona un manejo uniforme de las diferentes clases de datos basados en el tiempo, maximizando la preservación de la información de formato nativo.
- **Klar**, corresponde a una librería que contiene funciones para la visualización y la clasificación de datos.
- **Class**, corresponde a una librería que contiene funciones de clasificación supervisada en este caso el K-NN.

- **E1071**, corresponde una librería para funciones de clasificación supervisada y no supervisada.

5.2.-Datos utilizados

Para la realización del trabajo hemos usado: precios al cierre de las acciones, variación de la acción con respecto a la semana anterior y el volumen de operaciones de ese día. Hemos decidido usar estos tres elementos ya que el precio de la acción es el elemento diferenciador a la hora de invertir. La variación también es muy útil a la hora de ver el rango de precios que es capaz de adoptar una acción, a mayor volatilidad mayor riesgo. Por último, el volumen es interesante conocerlo ya que el precio suele fluctuar según el volumen de negociaciones de una acción.

5.2.1.-Precio de las acciones

Normalmente el tiempo en vivo de las acciones no está disponible en internet así que tenemos que trabajar con el histórico de la acción, la obtenemos en RStudio mediante la librería Quantmod, que muestra la información según la tabla 4.

Tabla 4.- Precios históricos de un valor

Date	Open	High	Low	Close	Adj close	Volume
2020-01-02	239.10	242.20	238.60	241.50	241.50	62001300
2020-01-03	241.20	243.30	239.10	239.40	239.40	42356212
2020-01-04	238.80	241.70	236.45	237.75	237.75	52005266
2020-01-05	237.90	240.10	237.30	239.70	239.70	64200550

Fuente: elaboración propia

La descripción es la siguiente:

- **Open:** precio de apertura de la acción, es decir, el precio de la primera transacción en el día.
- **High:** Cotización más alta a la que se comercializó la acción en ese día.
- **Low:** Cotización más baja a la que se comercializó la acción en ese día.
- **Close:** Precio de cierre, es decir, cotización de la última transacción del día.
- **Adj Close:** Precio de cierre ajustado a la división de las acciones.
- **Volume:** Cantidad de operaciones que se realizaron ese día.

5.2.2.-Obtención de los datos

Como se ha mencionado anteriormente vamos a trabajar con 34 empresas del IBEX-35 entre el 01-01-2020 al 10-03-2021. Diferenciando entre dos tipos de datos:

- **Datos de entrenamiento:** estos datos son los recogidos durante el periodo 01-01-2020 hasta el 31-12-2020. Los siguientes datos forman una matriz que es usada para entrenar el algoritmo y son datos que ya conocemos su predicción.
- **Datos de prueba:** son los datos recogidos entre el 01-09-2020 y el 10-03-2021. Estos datos son los encargados de hacer de predictivos y son con los que vamos a sacar conclusiones sobre la efectividad del algoritmo.

Para la realización de las matrices de entrenamientos nos hemos basado ligeramente en la cadena de valor de Porter(1980) como evidencia la figura 9, hace referencia a una serie de procesos de transformación que debemos seguir para generar datos de valor. Aunque sea un término que se aplica al Big Data seguir esta filosofía ayuda a estructurar mejor el recabado de información y el posterior análisis como indican Bhadani, A. K., & Jothimani en su libro[8]:

Figura 9.- Cadena de valor de Porter



Fuente: Adata de Bhadani y Jothimani (2014)

- 1) **Generación de datos:** este sería el primer paso, la creación de los datos que crearían nuestros datasets con los cuales realizamos nuestro análisis. En este caso serían las cotizaciones históricas de la bolsa.
- 2) **Recolección de los datos:** en este punto nos encargamos de recoger los datos que más nos interesan con diferentes herramientas o de forma manual.
- 3) **Transmisión de datos:** una vez se recolectan los datos hay que trasladarlos al lugar donde van a ser tratados. Nosotros en este caso no lo usamos.
- 4) **Preprocesamiento de los datos:** los datos obtenidos no siempre vienen en limpio por así decirlo y es preciso tratarlos para eliminar repeticiones, inconsistencias, etc. Este proceso se realiza con la intención de solo quedarnos

con los datos realmente útiles. En nuestro caso de los históricos quedarnos con la información más relevante.

- 5) **Almacenamiento de datos:** sería una plataforma fiable y de acceso fácil, que además sea capaz de soportar las cantidades de datos con los que se trabaja. En este ejemplo lo podríamos atribuir a las matrices ya que son como nuestro almacén de datos de fácil acceso.
- 6) **Análisis de los datos:** En este apartado ya con los datos limpios y bien estructurados podríamos analizar nuestros datos. En nuestro caso sería la implementación del algoritmo K-NN.

Tras conocer nuestro modus operandi, la intención es formar dos matrices la matriz train y la matriz test. La primera de ellas tendrá 12 columnas, esta matriz nos servirá para entrenar el algoritmo y mejorar las predicciones. La segunda será nuestra matriz test con 11 columnas que tendrá parte de los datos en común con la train para poder comprobar la eficacia del algoritmo la otra parte será a predecir. Ambas matrices se calculan igual excepto la última columna, esta columna es la que contiene las predicciones, que solo estará disponible en la de entrenamiento dado que en la matriz test es lo que queremos predecir.

En primer lugar, tras descargar el histórico de las cotizaciones introducimos el precio de cierre de las 34 acciones que nos interesan en un dataframe. Posteriormente con esos precios de cierre a partir de 60 días vamos obteniendo el rendimiento semanal de las acciones de forma repetitiva, es decir, de martes a martes, de miércoles a miércoles y así sucesivamente, durante 5 semanas atrás. De esta forma conseguiríamos nuestra primera matriz llamada matriz A. La fórmula del rendimiento es la siguiente:

$$\text{Rendimiento} = \frac{\text{Precio de venta} - \text{Precio de compra}}{\text{Precio de compra}}$$

Tras obtener la matriz de rendimientos pasamos a la de varianzas donde debemos coger esos días que agrupan esas semanas y calcular su varianza para poder comprobar su volatilidad y cómo se comportan. De esta forma conseguiríamos nuestra segunda matriz llamada matriz B. La fórmula de la varianza es la siguiente:

$$\text{Varianza} = \frac{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot f_i}{N}$$

Métodos de Inteligencia Artificial aplicados al comportamiento de acciones en bolsa

Para completar el resto de la matriz nos faltaría la columna de volumen, se obtiene directamente cogiendo el volumen del histórico. Y la columna de si sube o baja la acción solo disponible en la matriz train, esta columna nos indica si la acción sube mediante un 1 o si de lo contrario baja mediante un 0.

El objetivo sería conocer como fluctúa la bolsa, es decir, la Variable V12 en función del resto de las variables, siendo estas V1,V2,V3,V4,V5,V6,V7,V8,V9,V10,V. A continuación, se comenta como se han obtenido las variables, cual es cada una de ellas y a que matriz pertenecen cada una de ellas.

Quedando la matriz train compuesta por las variables de rendimiento V1, V2, V3, V4, V5. Dichas variables son las indicadas de representar el precio o más bien el rendimiento ya que se usa la comparación entre las diferentes semanas para obtener dicho rendimiento. Siendo más específico la V1 fue calculada teniendo en cuenta comparando el rendimiento con una semana anterior, es decir, se compararía el lunes actual con el lunes anterior. La V2 teniendo en cuenta el rendimiento del activo con respecto a las 2 semanas anteriores. La V3 con respecto a 3 semanas anteriores. Y por último la V4 y V5 con 4 y 5 semanas anteriores respectivamente.

Las variables de variabilidad en el precio son ,V6, V7, V8, V9, V10. Estas variables se encargan de reflejar la volatilidad del mercado. Estas variables se calculan usando la fórmula de la varianza que ya hemos comentado anteriormente. Funciona de forma similar a como se calculan las variables de rendimiento. Siendo la V6 calculada su varianza con respecto a la semana anterior, en este caso para ver las varianzas habría que agrupar los datos en semanas y calcular su varianza para comprobar su variación, es decir, en este caso cogeríamos los datos de lunes a viernes y sacaríamos la varianza de estos días en este caso sería la varianza de la anterior. La V7 se calcularía con respecto a las 2 semanas anteriores. Las variables V8, V9 y V10 se calcularían con respecto a las semanas 3, 4 y 5 semanas anteriores respectivamente.

Después tenemos la variable de volumen V, esta variable es la encargada de cuantificar la cantidad de transacciones en ese día. Dicha variable se calcula simplemente añadiendo la columna de volumen del histórico a nuestra matriz. Por último, la variable de resultados conocidos V12 está variable nos indica el resultado que da el algoritmo para poder entrenarlo. Dicha variable se ha obtenido comparando los rendimientos de las acciones y viendo si estos rendimientos eran superiores o inferiores. Si el

rendimiento era superior al anterior se le asignaba un 1 y por el contrario era inferior se le asignaba un 0. Por consiguiente, la matriz test quedaría con las mismas variables y funciones excepto la V12 ya que estas variables son las que vamos a usar para la predicción. La matriz test quedaría de la siguiente forma, V1, V2, V3, V4, V5, V6, V7, V8, V9, V10 y V.

5.3.-Funcionamiento

Tras la obtención de las matrices train y test ya podemos pasar a predecir con el primer método mediante la función K-NN de la librería `class`. La siguiente función, calcula las distancias mediante la distancia euclidiana y es lo que llamamos una caja negra, es decir, no sabemos cómo está programada en su totalidad. La función pide los siguientes argumentos:

- **Train:** matriz o data frame de entrenamiento
- **Test:** matriz o data frame de predicción. El vector tiene que poder ser interpretado como una columna de un vector o un caso.
- **Cl:** es el vector de clasificación de la matriz de entrenamiento.
- **K:** número de vecinos que vamos a usar.
- **Prob:** si la probabilidad es cierta será devuelta con el resultado.

Tras completar todos los argumentos obtenemos una serie de resultados diferentes, es decir, para cada K obtenemos un resultado. En nuestro caso hemos usado K=1, K=3, K=5, K=7. Siendo el K=3 el que mejor resultados nos ha aportado durante el uso del algoritmo.

El segundo método usaremos el método del Support Vector Machine mediante la función SVM de la librería `e1071`. El siguiente método es parecido al anterior, pero con alguna variante. Para empezar, hay que crear un modelo igual al anterior, pero en este caso usamos la función SVM en vez de la GLM, también en el modelo usamos el argumento `'kernel=linear'` que quiere decir que vamos a convertir el modelo en un modelo de Support Vector Machine siendo este más eficiente para estos casos. Tras generar el modelo, lo usamos para predecir los datos test mediante la función `predict`, por último, los comparamos con los datos ya predichos para comprobar su precisión.

Para el tercero de nuestros métodos vamos a usar la Regresión Logística, con este método mediante la función GLM creamos un modelo en función de la variable

Métodos de Inteligencia Artificial aplicados al comportamiento de acciones en bolsa

categoría V12 que es la variable que nos dice si sube o baja la acción en los datos de entrenamiento. La variable categórica es explicada por el resto de las variables quedando el modelo de la siguiente manera:

$$V12 \sim V1+V2+V3+V4+V5+V6+V7+V8+V9+V10+V$$

Después de generar el modelo usamos los coeficientes de las betas para multiplicarlos por las variables de la matriz según cada valor de la bolsa y se aplica una fórmula que por ejemplo sería:

$$1/1+\exp(-(-0.0291+0.3097x_1-2.2626x_2))$$

Generando el algoritmo un valor por debajo de 0.5, la cotización de la acción bajaría y siendo superior a 0.5, pronosticaría que la cotización de la acción va a subir.

6.- Resultados

Tanto el SVM como el K-NN han tenido una precisión del 73% y 76% respectivamente. Siendo la precisión de la Regresión Logística del 53% bastante inferior a los otros métodos. Como ilustran las tablas 6, 7 y 8 que se verán a continuación.

Como hemos mencionado anteriormente el $k=3$ ha sido el que mejor resultados nos ha aportado en el método K-NN con una precisión del 76%. Es una precisión más que aceptable ya que se sitúa en la media de los demás algoritmos que usan este tipo de tecnología. El SVM también se sitúa cerca de esa marca siendo una buena alternativa. Por último, la RL ha tenido problemas para detectar sobre todo los casos en los que las acciones bajaban ya que situaba casi todas las sesiones al valor por encima de 0.5, es decir, que la acción sube. Los datos con los que hemos obtenido dichos resultados han sido los correspondientes a las cotizaciones entre el 01-09-2020 y el 31-12-2020 ya que es donde se cruzan los datos de entrenamiento con los datos predictivos. Los resultados los hemos obtenido mediante una tabla de confusión. En particular, sirve para mostrar de forma explícita cuándo una clase es confundida con otra, lo cual nos permite trabajar de forma separada con distintos tipos de error. Se trata de una muy buena opción para clasificadores binarios donde se comparan los valores que se predicen con los reales. Se entiende TP por verdadero positivo y FP por falso positivo, de igual forma para los negativos como muestra la tabla 5.

Tabla 5.- Tabla de confusión

		Predicción	
		Sube	Baja
Real	Sube	TP	FN
	Baja	FP	TN

Fuente: elaboración propia

A continuación podemos ver la precisión de nuestros métodos mediante la tabla 6, tabla 7 y por último, la tabla 8:

Tabla 6.-Tabla de confusión K-NN

	BAJA	SUBE
BAJA	788	252
SUBE	244	745
Precisión		76%

Fuente: elaboración propia

Tabla 7.-Tabla de confusión SVM

	BAJA	SUBE
BAJA	766	271
SUBE	283	717
Precisión		73%

Fuente: elaboración propia

Tabla 8.- Tabla de confusión RL

	BAJA	SUBE
BAJA	79	952
SUBE	1	1005
Precisión		53%

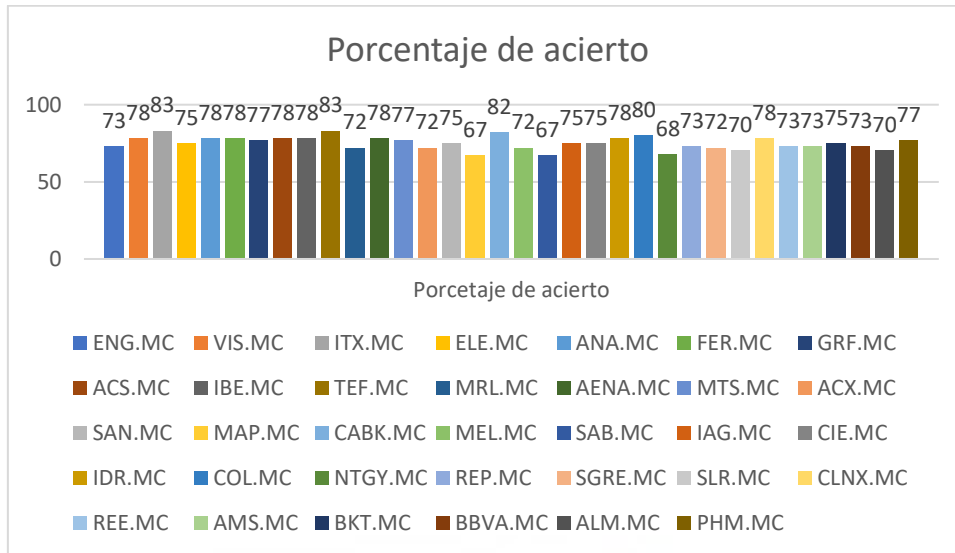
Fuente: elaboración propia

Teniendo en cuenta la prueba con las 34 empresas que cotizan en el IBEX-35 actualmente, recordemos que hemos excluido a Bankia. Por otra parte, fijándonos en las acciones de forma individual, las que mejor se han comportado con el método K-NN han sido Telefónica con una precisión del 83% e Inditex con la misma precisión que la

Métodos de Inteligencia Artificial aplicados al comportamiento de acciones en bolsa

anterior. Las que peor, Sabadell con una precisión del 67% y Mapfre con 67% también. Siendo la precisión media del método el 76%. Como se puede apreciar en la figura 10 donde también se encuentran el resto de los valores:

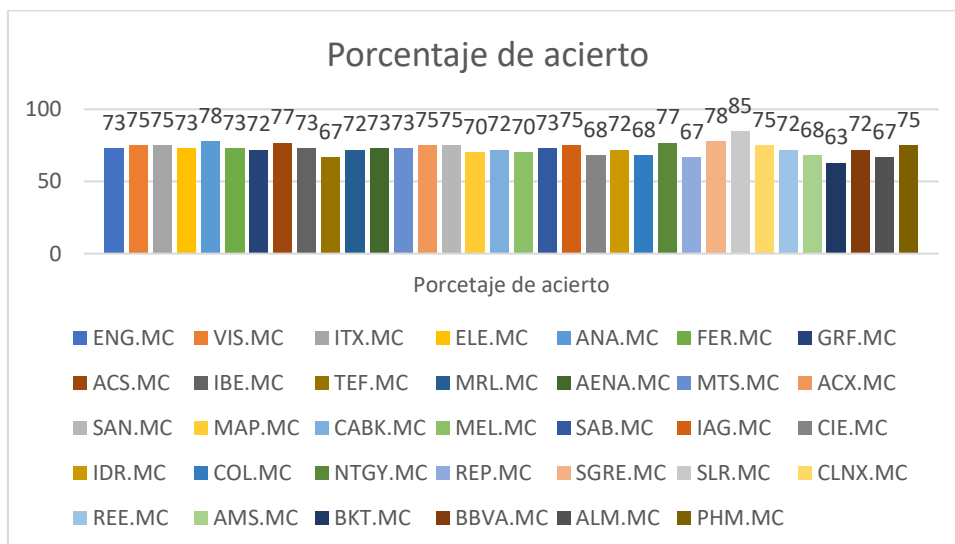
Figura 10.- Gráfico de barras de los porcentajes de acierto del método K-NN



Fuente: elaboración propia

En cuanto al método de Support Vector Machine su media se sitúa en un 73% de precisión, siendo el segundo mejor en este sentido y aplicable en la operativa diaria de un trader. El valor que mejor se han comportado ha sido Actividades de Construcción y Servicios (ACS) con una precisión del 77%. Por el contrario, tenemos a Bankinter (BKT) con una precisión del 63%. En la figura 11 representada a continuación se pueden ver todos los valores introducidos en el algoritmo.

Figura 11.- Gráfico de barras de los porcentajes de acierto del método Support Vector Machine

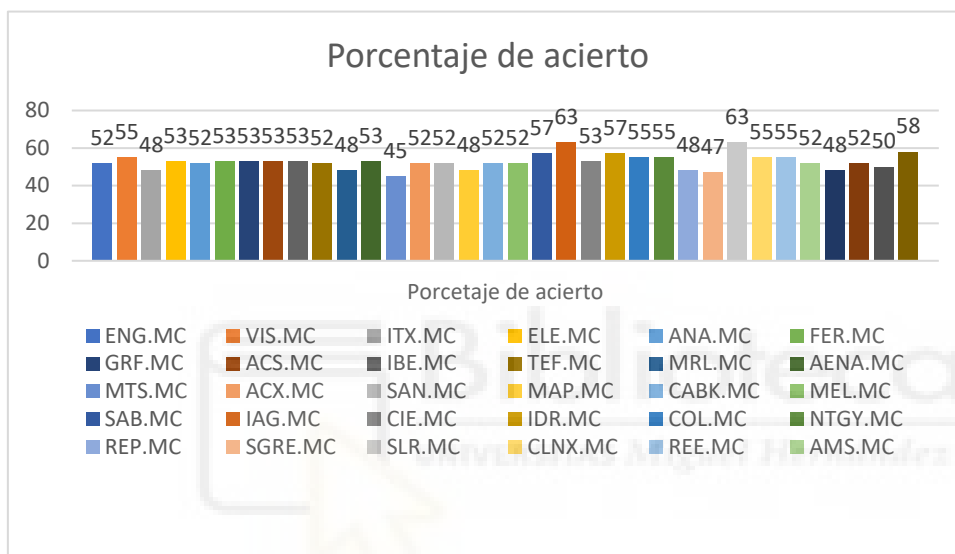


Métodos de Inteligencia Artificial aplicados al comportamiento de acciones en bolsa

Fuente: elaboración propia

El método de Regresión Logística ha funcionado de forma inferior al resto situando su media en torno al 53%. Siendo los valores que mejor se comportan International Airlines Group (IAG) con una precisión del 63%. El peor ha sido ArcelorMittal (MTS) con una precisión del 45%. La precisión de este método ha sido muy baja para poder usarlo de forma segura. En la figura 12 que se muestra a continuación se pueden ver las precisiones de todos los valores.

Figura 12.- Gráfico de barras de los porcentajes de acierto del método Regresión Logística



Fuente: elaboración propia

Tras la visualización de los resultados y tras la exposición de los métodos de nuestro algoritmo nos quedaríamos con el algoritmo compuesto por el K-NN ya que es el más preciso, por lo tanto, el que menor riesgo nos aporta en nuestra operativa bursátil. También cabe destacar de este método su facilidad de poner en práctica y su relación con la correlación de las acciones ya que se basa en comportamientos de acciones cercanas para discernir su tendencia.

7.-Conclusiones

El hecho de haber realizado este proyecto ha sido bastante productivo ya que se han conseguido los objetivos planteados al inicio de este, también se han conseguido otros objetivos secundarios y por supuesto el haber ahondado en el conocimiento sobre la

materia. En las siguientes secciones abordaremos las conclusiones extraídas del proyecto y las posibles mejoras en un trabajo futuro.

El uso de la Inteligencia Artificial en todas sus características ha quedado demostrado que los algoritmos de tipo predictivo pasarán a ser la piedra angular del sector de las inversiones financieras y se regularizará su uso en grandes compañías. Estos algoritmos producen grandes beneficios a la hora de manejar los datos producidos por las compañías, pero también es importante conocer los riesgos que entrañan y limitaciones para poder extraer su potencial de manera adecuada. Además, se ha podido demostrar, que pese a las componentes aleatorias que pueden existir en el comportamiento bursátil, hemos logrado predecir distintos movimientos y la implementación de modelos estadísticos como de inteligencia artificial en este caso con los métodos de K-Nearest Neighbour, Regresión Logística y Support Vector Machine. Estos métodos mejoran las posibilidades de obtener una inversión con menor riesgo y dado este factor un mayor rendimiento.

Por otra parte, el haber estructurado los datos en forma de rendimientos y variaciones ha sido una buena forma de probar un método diferente que no se centre únicamente en el precio, ya que el análisis técnico se centra en esta parte y explorar otras formas de inversión siempre es satisfactorio.

También la efectividad del algoritmo ha sido comprobada ya que según Juan M. Almodóvar[2] un algoritmo con una precisión superior al 55% ya puede ser usado. Esas decimas son bastante importantes, porque cada decima podría ser o una ganancia más o una pérdida menos según se mire.

8.- Trabajo futuro

Como algoritmo esto sería solo el principio ya que es un algoritmo que trabaja con datos históricos, es decir, datos que están almacenados una vez han pasado. Lo ideal sería trabajar con datos en tiempo real e implementar herramientas con la que podamos conseguir un flujo continuo de datos y trabajar con otras variables como pudieran ser, las correlaciones y como afecta el movimiento de cierta acción a nuestros intereses, el entorno de la acción también sería importante de poder cuantificarlo ya que todas las acciones por su sector no tienen el mismo riesgo.

También otra posible mejora, sería poder implementar otros métodos de Machine Learning como el discriminante lineal, o ya con métodos más avanzados de Inteligencia Artificial como redes neuronales. Estos algoritmos tal vez hagan subir unas décimas la precisión y mejoren el rendimiento. Otra de las posibles mejoras es la introducción de nuevas variables que aporten más información al modelo y mejoren su predicción

9.-Bibliografía

- [1] Formación BBVA. *Factores que influyen en el precio*.
- [2] Rankia. (2015). *Machine Learning aplicado al trading*.
www.rankia.com/blog/juan-almodovar/2623448-machine-learning-aplicado-trading
- [3] Acelera.tech. *Predecir el valor de una acción usando series de tiempo*.
acelera.tech/2020/11/26/machine-learning-predecir-el-valor-de-una-accion-usando-series-de-tiempo/
- [4] Rankia. Esteban Pérez (2019). *Trading Algorítmico con redes neuronales*. www.rankia.com/blog/como-no-quedarte-a-2-velas-en-el-trading
- [5] Lasse Rouhiainen (2018). *Inteligencia Artificial 101 cosas que debes saber sobre nuestro futuro*. Alienta editorial.
- [6] Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani (2013). *An Introduction to statistical Learning*. Springer
- [7] Francisca Serrano (2019). *Análisis técnico de bolsa y trading*. Dummies
- [8] Bhadani, A. K., & Jothimani (2016). *Big data: challenges, opportunities, and realities*.
- [9] Dash, R., & Dash, P. K. (2016). *A hybrid stock trading framework integrating technical analysis with machine learning techniques*.
- [10] De Mauro, A., Greco, M., & Grimaldi, M. (2015). *What is big data? A consensual definition and a review of key research topics*.

[11] Blog del Instituto de ingeniería del conocimiento (2017). *Big data en bolsa una inversión segura*. <https://www.iic.uam.es/innovacion/big-data-en-bolsa/>

[12] Blog slowinver.com (2013). *Algoritmo bandada para predecir la bolsa*. slowinver.com/algoritmo-bandada/

[13] Apuntes tema 7 Universidad Carlos III de Madrid (UC3M) (2012) *Introducción a las series temporales*.

[14] Charu C. Aggarwal (2018). *Neural Network and Deep Learning*. Springer

[15] Sudeep Tanwar, Sudhansu Tyagi Neeraj Kumar (2020). *Multimedia Big data Computing for Iot Applications*. Springer

[16] Blog Bolsa y Mercados Españoles.

www.bolsasymercados.es/esp/Sobre-BME/Indices

[17] Apuntes de técnicas estadísticas en análisis de mercados, José Luis Sainz-Pardo Auñón. *4.1 K-Nearest Neighbour*.

[18] Apuntes de técnicas estadísticas en análisis de mercados, José Luis Sainz-Pardo Auñón. *4.3 Support Vector Machine*.

[19] Apuntes de técnicas estadísticas en análisis de mercados, José Luis Sainz-Pardo Auñón. *4.4 Regresión Logística*.

[20] Abdelmalik Moujahid, Iñaki Inza y Pedro Larrañaga (2010). *Clasificadores K-NN*.

[21] Comprehensive R archive network. cran.r-project.org/packages

[22] Blog Portfolio. *Predecir el futuro de los mercados financieros ¿Realidad o mito?*

10.-Anexo

