



Facultad de Ciencias Económicas y Empresariales, ICADE

Machine Learning aplicado al trading: marco teórico y análisis de casos prácticos

Clave: 201501427

“All models are wrong, but some are useful.”

— (George Box, 1976)

“Sin alcanzar la certeza, la estadística nos ofrece, al menos, una medida de la incertidumbre”

— (López y Queralt, 2019)

Para Elena, porque las matemáticas son nuestro futuro.

Índice

1. Introducción	4
a. Objetivo	4
b. Justificación	4
c. Metodología	6
I. Enfoque deductivo	6
i. Ventajas del método deductivo	6
ii. Fuentes utilizadas	7
II. Herramientas de análisis	7
d. Estructura	7
2. Machine Learning	8
a. Concepto	8
b. ¿Cómo funciona el machine learning?	10
c. Tipos	13
d. Metodología	15
I. Dataset	15
II. <i>Feature engineering</i> o ingeniería de las características	16
III. Entrenamiento de nuestros datos	19
IV. Evaluación del modelo	22
V. Otras consideraciones	23
i. Sobreajuste	23
ii. Infraajuste	23
VI. Selección del Modelo	24
e. Aplicaciones	24
3. Trading	27
a. Principios básicos	28
b. Aplicaciones de estadísticas a la inversión	29
I. Análisis de valores	30
II. La teoría del Dow	32
III. Análisis técnico-estadístico	32
IV. La nueva revolución tecnológica	36
c. Trading algorítmico	36

II.	Flash trading.....	37
III.	High Frequency trading (HFT)	37
4.	<i>Aplicación del machine learning al trading: Ejemplos prácticos</i>	40
a.	¿Qué modelo utilizar?	40
b.	Supervisado: Clasificación	41
c.	No Supervisado: Clusterización	43
d.	Caso práctico	47
I.	Reducción de dimensionalidad: Clustering Jerárquico	47
II.	Clusterización particional	48
III.	Regresión.....	52
e.	Interpretación general de los resultados obtenidos	57
5.	<i>Conclusiones</i>	59
6.	<i>Bibliografía</i>	61
7.	<i>Anexos</i>	63
a.	R code	63
I.	Webscraping de las empresas del IBEX35	63
II.	Modelo ARIMA	64
III.	Clustering con DTW y Dendograma.....	65
IV.	Charting del oro.....	67
b.	Python code	67

Machine Learning aplicado al trading: marco teórico y análisis de casos prácticos

1. Introducción

a. Objetivo

El propósito de este trabajo es estudiar el potencial del machine learning para desarrollarlo en los mercados financieros, y como ejemplo de ello tomaremos el trading como campo particular del mundo empresarial y las finanzas, partiendo de la base de que ningún aspecto de nuestra vida cotidiana, ni de nuestra economía es ajeno a la incipiente revolución tecnológica que nace de la ciencia de datos. Esta revolución es un hecho y de su adaptación al entorno empresarial dependerá la viabilidad a largo plazo de la proposición de valor de cualquier modelo de negocio.

De hecho, en los mercados financieros encontramos que el uso de los algoritmos, la Inteligencia Artificial y dentro de esta el machine learning, son una realidad incipiente que prometen revolucionar su tradicional *modus operandi*.

En definitiva, ante un cambio disruptivo de carácter predominantemente tecnológico, social y económico; este trabajo tiene un fin eminentemente exploratorio, para tratar de aportar valor a un aspecto fundamental de las economías de mercado por su capacidad de aportar financiación y liquidez: el trading.

El objetivo concreto es: analizar y elaborar modelos de machine learning aplicables al trading. Partiendo de modelos analizados en la literatura trato de replicarlos en un caso práctico.

b. Justificación

La evolución de la capacidad de computación, tanto de almacenaje como de procesamiento de datos ha permitido el impulso de una nueva revolución tecnológica, que abre las puertas a una nueva posibilidad de comprensión del entorno. Las nuevas tecnologías son el motor de una incipiente revolución. Aquí algunos datos que avalan los avances revolucionarios en la computación:

- En 1980 un ordenador IBM de 20 GB de RAM costaba un millón de dólares y pesaba dos toneladas, hoy una microSD alberga 36 GB pesa medio gramo y cuesta menos de diez dólares.
- En 1971 el primer microprocesador Intel trabajaba a 700 kHz, hoy un microprocesador Intel i7, muy popularmente encontrados en ordenadores actuales de gama media alta,

trabaja hasta 4,5 GHz a una frecuencia más de 6400 veces mayor¹. Ese mismo microprocesador trabaja a 33.120 MIPS (millones de instrucciones por segundo). El primer coche autónomo desarrollado en Stanford en 1979 funcionaba con un procesador de tan solo 2,5 MIPS.

- La información generada en los últimos dos años de la humanidad equivale a toda la información creada, registrada y almacenada durante toda la historia de la humanidad anterior a esos dos años². Y la tendencia crece exponencialmente.

Nos encontramos ante un cambio disruptivo: tenemos acceso a cantidades ingentes de información, encerrada en los datos; pero los datos, por sí solos, no son capaces de brindarnos autónomamente el conocimiento que los hacen realmente útiles.

Para ello, necesitamos ser capaces de sintetizarlos, relacionarlos y contextualizarlos a fin de comprender los fragmentos de realidad que encierran, la verdad secreta que ocultan a primera vista.

El Big Data supone el punto de partida de esta nueva revolución tecnológica. La definición de Big Data no es más que el manejo de una gran cantidad de datos, de información, siempre que cumpla cuatro características: volumen, variedad, veracidad y velocidad³. Posibilidad que nos brinda los avances en el campo de la ciencia de la computación, en especial en cuanto la capacidad de procesamiento y de almacenaje⁴, y más importantemente la accesibilidad a esta tecnología en términos económicos.

Gracias a las técnicas de Big Data somos capaces de generar, guardar, conectar, acceder y entender más información y sobre todo sintetizarla. El análisis masivo de datos es realmente un cambio disruptivo en nuestra sociedad y sus aplicaciones están aun en pleno auge, pese a que su potencial real es aún desconocido. A la par del desarrollo del Big Data surgen a su vez nuevos avances en las técnicas de análisis de datos.

El Big Data nos permite establecer conexiones entre información, entenderla como un conjunto, agruparla o fragmentarla mediante procesos automatizados, métricas, análisis, gráficas, etc. Las técnicas que abarcan el Big Data son, del mismo modo aun, una ciencia en desarrollo. El Big Data alberga la posibilidad de acceder a un nuevo entendimiento al que antes teníamos cerradas las puertas. El mundo es capaz ahora de conocer más allá de los límites

¹ Ibíd p. 41

² Ibíd p. 41

³ Ibíd p. 41

⁴ Ibíd p. 12

naturales de nuestra capacidad cognitiva, y el mundo empresarial no es ajeno a este hecho, y ya se hace eco de esta tecnología para transformarla en valor.

Dentro de estas diferentes técnicas este trabajo se va a centrar en las aplicaciones de una concreta metodología análisis de datos derivada de la inteligencia artificial, el machine learning. La Inteligencia Artificial (IA) es un concepto en sí amplio que recoge todas aquellas tareas intelectuales que tiene como objetivo sustituir la intervención humana por la de una máquina. Una definición ampliamente reconocida de la IA es:

“el esfuerzo por automatizar tareas intelectuales normalmente realizadas por humanos”⁵

Como anticipo, el machine learning supone una nueva herramienta con la que desarrollar modelos matemáticos o estadísticos, comúnmente se divide en varios tipos: predictivos, clasificatorios o de clustering; o bien supervisados o no; de los que podemos desprender una relación matemática de los distintos datos objeto de estudio, y nuestras variables, de manera que nuestra herramienta sea capaz de aprender por sí misma.

c. Metodología

Para llegar a una conclusión sobre nuestro objetivo de investigación vamos a seguir un razonamiento deductivo.

I. Enfoque deductivo

Partiendo de la premisa de que el machine learning puede crear valor en el trading, vamos a construir un ejemplo práctico, *ad hoc*, para determinar si esta premisa es falsa. Si los resultados no son concluyentes o son contrarios a la premisa automáticamente tenemos que descartar su aplicabilidad como principio genérico.

i. Ventajas del método deductivo

Si se utilizase el método inductivo únicamente podríamos afirmar la validez de la premisa para el caso concreto estudiado. Sin embargo, utilizando el método deductivo podemos con una sola observación comprobar si es falsa, si la premisa es cierta es aplicable a todos los casos concretos. Y puesto que nos interesa más conocer si es genéricamente falsa que específicamente verdadera, seguiremos este enfoque.

⁵ (Chollet, 2019) p. 24

ii. Fuentes utilizadas

- Primarias:
 - Yahoo finance
 - Oanda
 - Leyes e informes de autoridades oficiales
- Secundarias
 - Recursos de internet
 - Revistas y publicaciones académicas

II. Herramientas de análisis

Las herramientas utilizadas para el análisis de datos son los lenguajes de programación R y Python, en los entornos RStudio y Google Colab, respectivamente; y Excel.

d. Estructura

Este trabajo se divide en cuatro grandes partes sustanciales, que abordan directamente el objetivo de investigación:

- La parte 2 de explicación del concepto y funcionamiento del machine learning y sus aplicaciones y usos comunes.
- La parte 3 de introducción al mundo del *trading* y la inversión desde una faceta regulatoria, económica y estratégica, haciendo alusión al encaje de la estadística y la automatización del sector por medio de algoritmos, particularmente.
- La parte 4 de estudio de casos prácticos de implementación del machine learning al *trading*.
- Y, por último, la parte 5 que recoge las conclusiones de la investigación de los potenciales usos, junto las perspectivas futuras del devenir del machine learning y la Inteligencia Artificial.

2. *Machine Learning*

¿En qué consiste el Machine Learning, y cómo puede suponer una herramienta disruptiva en el trading, qué nuevas posibilidades concretas ofrece que aporten valor a los traders?

“El machine learning es un subcampo de las ciencias de la computación que se ocupa de construir algoritmos que, para ser útiles, se basan en una colección de ejemplos de algún fenómeno. Estos ejemplos pueden provenir de la naturaleza, ser hechos a mano por los humanos o generados por otro algoritmo.”⁶

Por lo tanto, el machine learning nos brinda la posibilidad de desarrollar una técnica que durante los últimos años está revolucionando los mercados financieros y las estrategias de inversión de los mayores fondos internacionales, el trading algorítmico, que a partir de una serie de variables ejecuta diferentes órdenes preprogramadas para optimizar la estrategia de inversión⁷.

Del trading algorítmico han surgido exitosas técnicas de inversión como el *flash trading* o el *high frequency trading* (HFT). Algunos de estos fondos cuánticos incluso han conseguido superar la rentabilidad media anual de exitosos fondos liderados por prestigiosísimos gurús de la inversión como Peter Lynch, Warren Buffet o George Soros, como veremos más adelante.

a. **Concepto**

El machine learning es una ramificación concreta de la inteligencia artificial. El término fue un concepto de marketing acuñado por Arthur Samuel, pionero de la inteligencia artificial y empleado de IBM en 1959 para describir una función de la que algunas computadoras eran capaces de imitar y sustituir tareas intelectuales tradicionalmente realizadas por humanos. Y es que, estos ordenadores eran capaces de realizar tareas para las que no estaban explícitamente programados⁸, ¿o sí?

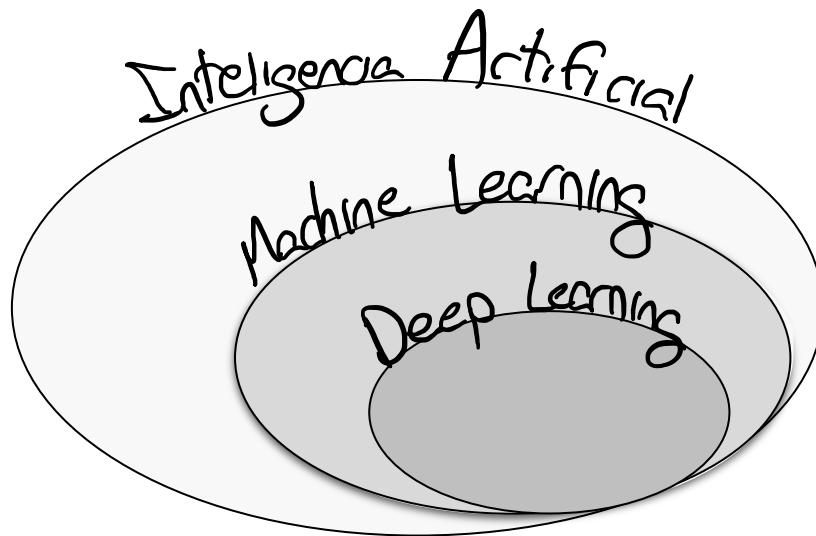
⁶ (Burkov, 2019) p. 3

⁷ [Algorithmic Trading](#). Visto el 20/09/2020 en:

<https://www.investopedia.com/terms/a/algorithmictrading.asp#:~:text=Algorithmic%20trading%20is%20a%20process,to%20the%20market%20over%20time>.

⁸ (Burkov, 2019) p. 3

Ilustración 1 Encaje del Machine learning en la IA, (Fuente: elaboración propia)



Consiste en el aprendizaje automatizado por parte de una máquina, el ordenador, en base a una serie de inputs: los datos. A partir de ellos nuestra máquina es capaz de encontrar relaciones entre los mismos: patrones.

El proceso de aprendizaje como tal no existe para la máquina, es un artificio. Lo que realmente hace la máquina es seleccionar una fórmula matemática concreta que, aplicada a nuestros inputs de entrenamiento, dentro de un set de datos, devuelve los outputs deseados por el modelizador. Con la capacidad añadida de que esa misma fórmula sería aplicable a otros datos distintos del conjunto de entrenamiento, siempre que los nuevos datos hayan sido extraídos de una misma o parecida distribución estadística a la que siguen los datos de entrenamiento.

Su utilidad es aprovechada principalmente para elaborar modelos predictivos de clasificación o regresión de un nuevo output, utilizando esa relación matemática que interrelaciona las variables de nuestros datos, el algoritmo, para realizar conjeturas sobre un nuevo conjunto de datos, siempre que estos sigan la misma distribución que los datos originales empleados para el desarrollo de algoritmo; o también para elaborar clústeres, particiones arbitrarias de nuestro conjuntos de datos en función de unas restricciones impuestas, normalmente en base a su cercanía (por ejemplo, en el *K-Nearest-Neighbours*).

El uso del machine learning supone modelizar, por lo que conviene recordar que cualquier modelo por mucho que se acerque a la realidad es falso por definición, aunque no por ello significa que sea inútil, o que no se ajuste a la realidad. Al menos en cierto grado.

En conclusión, el objetivo del machine learning es idéntico al de cualquier otra técnica de modelización:

Hacer una aproximación lo más cercana a la realidad posible, optimizando el error y la precisión de nuestro modelo.

Para ello se utilizarán distintas métricas como criterio diferenciador. Pero entonces, ¿por qué usar un modelo si es incapaz de ajustarse completamente a la realidad? Pues por la incapacidad de establecer una relación matemática entre nuestros datos, modelizar es la única manera de obtener una aproximación a esa relación. Existen situaciones en las que, por su complejidad al intervenir en ella un gran número de factores, o por el desconocimiento total o parcial de las causas del fenómeno estudiado, la mejor manera de intentar entenderlas supone construir un modelo.

En definitiva, en base a información histórica y mediante un proceso automatizado nos permite establecer una relación entre inputs y outputs en un lenguaje universal que podamos entender nosotros y los ordenadores: las matemáticas. Siempre que cumpla las necesarias condiciones estadísticas, particularmente en cuanto a la mencionada distribución de nuestros datos de entrenamiento.

b. ¿Cómo funciona el machine learning?

El funcionamiento del machine learning parte de una idea simple pero tremendamente útil. Si en un programa informático clásico funciona de la siguiente manera: el programador introduce unos datos, después el programa los procesa según las restricciones impuestas por el programador, y en base a aquellas limitaciones el programa devuelve el output correspondiente a los datos introducidos; el machine learning reordena las fases de este esquema.

Su funcionamiento parte de la siguiente manera: primero se introducen los datos, de forma que a cada input le corresponda un output en base a una posible relación que el modelizador desea conocer.

El modelizador desconoce las reglas de dicha relación y para conocerlas, o más bien cuantificarlas, entrega la información a la máquina (nuestro ordenador) de los respectivos pares de inputs con sus outputs relativos a cada observación de nuestra muestra, de manera que la máquina es capaz de deducir una relación, que se perfecciona con cada nueva observación introducida. Es precisamente en el perfeccionamiento de la relación donde hablamos del aprendizaje de la máquina. Pero ¿Cómo es capaz la máquina de aprender, es decir, de establecer esa relación en base a la nueva información que se le introduce?

Precisamente mediante un programa. La idea de la que parte el machine learning es la de optimización. Nuestra máquina es capaz de elegir el algoritmo que mejor se ajusta a los datos

escogiendo aquella fórmula matemática que los relaciona, minimizando el error, medido según diferentes métricas dependiendo de las características de nuestro modelo.

Gracias a la capacidad de computación el ordenador ejecuta los posibles algoritmos que relacionan nuestras variables, escogiendo al más preciso de estos algoritmos, entendiendo este como el que menos falle, en términos relativos.

En resumen, los algoritmos de aprendizaje se construyen sobre tres cimientos:

- **Una función de pérdida:** normalmente penaliza el error del modelo en base a predicciones incorrectas. Una función clásica de pérdida es el RMSE (*Root Mean Squared Error*)⁹ o en una regresión lineal el error cuadrático.
- **Un criterio de optimización:** ha de minimizar el error de la función de pérdida. Comúnmente si la función de pérdida es diferenciable se utiliza la optimización por gradientes: el gradiente descendiente y particularmente el gradiente descendiente estocástico. Básicamente derivando la función de pérdida para alcanzar los valores de los parámetros de nuestro modelo que minimizan el error. Moviéndose en la dirección opuesta a la del gradiente como veremos.
- **Optimización usando el dataset**¹⁰: un método para encontrar el modelo óptimo usando nuestros datos. Computacionalmente esto se logra en base a iteraciones sobre las observaciones que actualizan la función óptima, según una tasa de aprendizaje α . Con cada nuevo ejemplo del conjunto de entrenamiento se afina el algoritmo.

Por ejemplo:

Para optimizar los parámetros de un algoritmo SVM (*Support Vector Machine*), algoritmo definido por los parámetros w y b , que devuelve una predicción y' :

$$SVM(w, b) = wx + b = y'$$

La función de pérdida toma la diferencia, la distancia entre los valores y' predichos y los valores objetivo y , utilizando el error cuadrático para simplificar su tratamiento, sería:

$$Error(w, b) = (y - y')^2 = (y - wx + b)^2$$

Manteniendo los pares I/O, de inputs y outputs¹¹, x e y , estáticos tratamos los parámetros w y b como variables y buscamos optimizarlas intentando obtener el menor error.

Calculamos los gradientes, utilizando la regla de la cadena:

⁹ Raíz cuadrada del error medio cuadrático.

¹⁰ Nuestro conjunto de datos.

¹¹ Comúnmente denotamos estos pares de valores como pares I/O.

$$\frac{\partial Error(w, b)}{\partial w} = -2x(y - (wx + b))$$

$$\frac{\partial Error(w, b)}{\partial b} = -2(y - (wx + b))$$

Ahora, estos gradientes son aplicables para pares I/O individuales. Pero podemos generalizar esta expresión para encontrar los valores w y b que minimicen el error de nuestro par x_i e y_i , y que compute la pérdida global del cambio de los parámetros, ajustando según el peso del par en el conjunto de entrenamiento (N):

$$\frac{\partial Error(w, b)}{\partial w} = \frac{1}{N} \sum_{i=1}^N -2x_i(y_i - (wx_i + b))$$

$$\frac{\partial Error(w, b)}{\partial b} = \frac{1}{N} \sum_{i=1}^N -2(y_i - (wx_i + b))$$

Con cada nuevo par I/O actualizamos los parámetros w , b fijados inicialmente (por ejemplo: $w = 0$, y $b = 0$), a la tasa α , y siempre en la dirección opuesta a la del gradiente, iterando por cada uno de los datos del conjunto de entrenamiento. Inicialmente se fijan valores de los parámetros aleatorios, para que se vayan actualizando con cada nueva entrada de datos, lo que entendemos como que nuestro modelo vaya aprendiendo. Tras varias *epochs*¹² nuestros parámetros tienden a fijarse entorno a unos valores:

$$w_i \leftarrow \frac{2x_i(y_i - (w_{i-1}x_i + b_{i-1}))}{N} \times \alpha$$

$$b_i \leftarrow \frac{2(y_i - (w_{i-1}x_i + b_{i-1}))}{N} \times \alpha$$

¿Por qué actualizamos en la dirección opuesta del gradiente? Porque si el gradiente nos indica cual es la pendiente de la función en un punto entonces:

- Si la pendiente es positiva, significa que la función en el siguiente punto mayor crecerá. Matemáticamente, en nuestra función error:

$$\text{Si: } Error'(w_i) > 0; \text{ entonces: } Error(w_{i+1}) > Error(w_i)$$

El error aumenta si aumenta el parámetro, entonces nos interesa mover nuestro parámetro tomando la dirección contraria, hacia w_{i-1} , esto es tomar la dirección opuesta a la del gradiente

- Si es negativa, por consiguiente:

¹² Cada vez que iteramos a través de todo el conjunto de entrenamiento. Dos *epochs* suponen iterar 2 veces por todos los pares del dataset.

Si: $Error'(w_i) < 0$; entonces: $Error(w_{i+1}) < Error(w_i)$

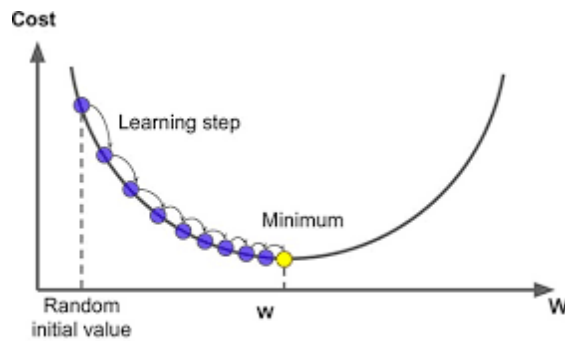
Aquí del mismo modo nos interesa mover nuestro parámetro en la dirección contraria, hacia w_{i+1} . Aplica la misma lógica.

Considerando que:

$$w_{i-1} < w_i < w_{i+1}$$

Visualmente podemos representar esto como una pelota que cae sobre una curva, y que acaba depositándose en el mínimo de la función:

Ilustración 2 Funcionamiento del Gradiente Descendiente (Fuente: <https://morioh.com/p/15c995420be6>)



c. Tipos

Tabla 1 Tipos de Machine Learning (Fuente: elaboración propia)

Según	Tipos de Machine Learning	
Cómo se construye el modelo	<i>Instance based</i>	<i>Model based</i>
Las capas que utiliza	<i>Shallow Learning</i>	<i>Deep Learning</i>
Método de Aprendizaje	Semi-supervidado	Reforzado
	No supervisado	Supervisado

		clasificación	regresión
estacionarios	series temporales (técnicas de reducción de dimensionalidad)		
particional	fuzzy		
jerárquico	crisp	jerárquico	
densidad			particional
<i>grid-based</i>			
<i>model-based</i>			

Dependiendo de diversos criterios podemos distinguir distintos tipos de machine learning, esquematizados en la Tabla 1. Comúnmente se distinguen¹³:

- Según el tipo de aprendizaje (más adelante estos serán detallados en el apartado de metodología):
 - Modelos supervisados
 - Modelos no supervisados
 - Semi-supervisados
 - Modelos reforzados
- Según se construya el modelo:
 - *Instance-based*: utilizan todo el dataset como modelo, ejemplo de ello es el algoritmo kNN (*K-Nearest Neighbors*) de clasificación, que clasifica las nuevas observaciones en base a la clase a la que pertenece la observación de nuestro modelo más cercana. Por ejemplo, utilizando distancias euclídeas.
 - Modelos paramétricos: son preconstruidos que se afinan con el conjunto de entrenamiento. Por ejemplo, el SVM (*Support Vector Machine*) o la regresión lineal, el machine learning escoge los parámetros que optimizan el error de estos modelos.
- Según las capas empleadas:
 - *Deep learning*: utiliza modelos anidados, en los que se construyen varias capas. Dependiendo de los outputs aprendidos de las primeras capas se aprenden a su vez los outputs finales. Se produce un aprendizaje sobre el aprendizaje. El ejemplo paradigmático de este algoritmo son las redes neuronales. Sin duda esta será en el futuro la vertiente con mayores posibilidades del machine learning.
 - *Shallow learning*: tan solo se utiliza una capa. Los outputs son directamente aprendidos del dataset de entrenamiento. El *deep learning* no es más que concatenaciones de *shallow learning*.

¹³ *Ibíd.* ps. 12-13

d. Metodología

En este apartado vamos a profundizar en el proceso técnico, genérico, que tiene que seguir un modelizador. Antes de nada, necesitamos trabajar con una información, teniendo siempre en cuenta que ha de ser información relevante, esto significa que debe aportar capacidad predictiva al modelo. Cuando establecemos sobre qué información vamos a trabajar, lo idóneo sería que exista una fuerte relación de causalidad.

Y bien, ¿en qué consiste el funcionamiento específico de esta técnica para la elaboración de un modelo estadístico usando machine learning?

I. Dataset

Inicialmente partimos de un set de datos, información que hemos de considerar que goza de capacidad explicativa, organizada en observaciones y variables, nuestras filas y columnas respectivamente, que queramos analizar.

Como requisito para la mayor validez de nuestro modelo, el dataset ha de ser extraído de una muestra, que gracias al Big Data puede ser enorme y asemejarse mucho a la población de la que se saca. La muestra ha de cumplir con los requisitos estadísticos que garanticen que es representativa de la población de estudio para poder extrapolar a la población las conclusiones a las que lleguemos, si es posible.

De entre todas las variables para intentar evitar el principal contra del machine learning, el **sobreajuste**, preseleccionaremos aquellas que puedan ostentar un mayor poder predictivo. La práctica habitual implica estudiar:

- La distribución individual de cada variable: distribuciones uniformes reflejan que la variable estudiada es aleatoria o cuasi aleatoria con lo cual no funciona como variable explicativa, no aporta información, y habrá que descartarla.
- La matriz de correlaciones: las correlaciones nos indican el porcentaje de variabilidad de una variable asociado a la variabilidad de otra, es decir si una cambia la otra también lo hará más cuanto mayor sea la correlación. Para elaborar nuestro modelo nos interesara incluir como variables explicativas aquellas más correlacionadas con nuestra variable endógena¹⁴.
- Planteamiento teórico: es posible que incluyamos variables que no cumplan con los requisitos anteriores cuando la intuición del investigador considere incluir determinadas variables por sospechar que puedan ser explicativas.

¹⁴ Aquella que queremos predecir utilizando las variables exógenas, estas son las explicativas.

II. *Feature engineering* o ingeniería de las características

Elaborada nuestra muestra organizamos las variables que componen el dataset con los datos muestrales. Hemos de diferenciar en este paso entre variables numéricas y categóricas, para mayor corrección de nuestro modelo, además de considerar si resulta conveniente el tratamiento de ciertas de variables de una manera específica, un aspecto realmente importante para conseguir un buen modelo. Este es proceso del *feature engineering*, en el que se busca sacar el mayor provecho de la información disponible. Existe un tipo de machine learning que por su complejidad nos permite omitir parcialmente este paso (salvo por la normalización y codificación numéricamente), este es el Deep learning, lo que supone una gran ventaja.

No existe per se un modus operandi de cómo tratar los datos, aquí juega un papel importante la habilidad del investigador para jugar con sus datos a fin de desarrollar el mejor modelo posible. Algunas técnicas comunes son:

- Logaritmos: utilizado cuando el rango de nuestra variable numérica es muy amplio y lo que nos interesa para nuestro modelo es realmente los cambios relativos y no absolutos. Por ejemplo, es habitual su uso en variables como la renta económica de una persona en la que realmente nos interesa saber diferenciar entre franjas o intervalos de renta similares, y no diferenciar en términos absolutos.
- Estandarización: cuando la distribución de nuestra variable numérica se asemeja a una normal es conveniente darle un tratamiento estadístico como tal, estandarizándola, para su mejor manejo. Otra circunstancia que resuelve la estandarización es la de comprimir variables con *outliers*. Debemos tener en cuenta que en la estandarización la distribución original de nuestra variable se sustituye por la de la normal, con todas sus propiedades. Con las consecuencias que ello implica

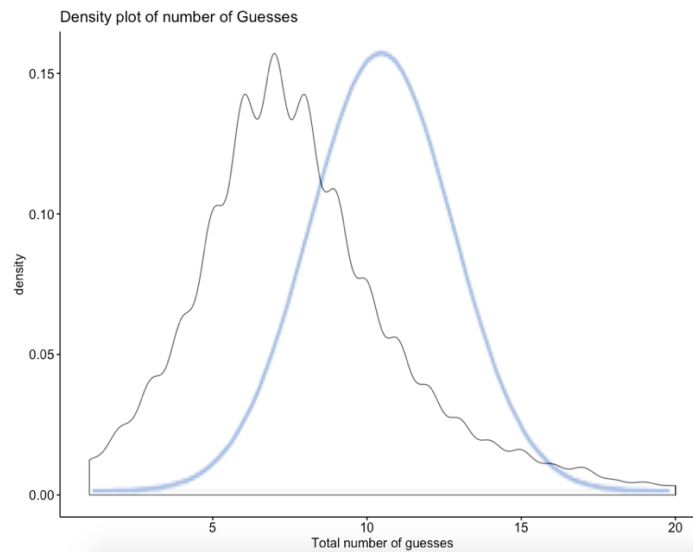
Dado que cualquier normal converge a su vez en una normal de media 0 y desviación típica 1, facilita la manipulación de los datos ya que, gracias a las tablas de distribución de probabilidad de la normal, conocemos a su vez la distribución y la probabilidad asociada a cualquier normal. Existen dos métodos por los que asumir que una distribución es normal:

- **Gráficamente**

Este método compara visualmente o bien:

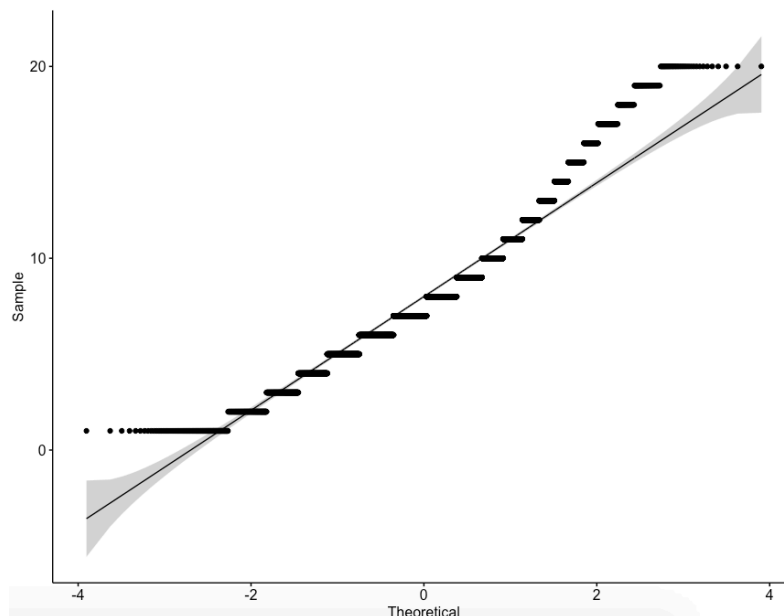
- la función de densidad de la variable y su similitud con una normal, o,

Gráfico 1 Ejemplo de función de densidad frente una normal (Fuente: elaboración propia en R)



- la gráfica *quantil-quantil*, que muestra los puntos sobre los que recae la distribución de nuestra variable (de color negro) en contraste con el área sobre la que debería recaer de asemejarse a una normal (de color gris).

Gráfico 2 Ejemplo de gráfica *quantil-quantil*, para la misma variable (Fuente: elaboración propia en R)



○ Estadísticamente

Existen varios estadísticos utilizados para comprobar la normalidad. Todos ellos trabajan bajo la hipótesis nula de normalidad, por lo que si por ejemplo fijamos un nivel de significación del 5% la normalidad se cumpliría de obtener un p-valor $> 5\%$.

Por simplicidad podemos trabajar con dos tests de normalidad:

- El test de Shapiro-Francia y,
- el test chi cuadrado de Pearson

Al ser los datos de cada variable los inputs que el machine learning utiliza para entrenar su modelo a fin de dar con el output correcto, cuantos más datos pasen por el aprendizaje automático y más información puedan aportar estos mejor será nuestro modelo. Es decir, el aprendizaje mejora cuanto más información absorba nuestro modelo.

Adicionalmente existen otras técnicas para lograr mejores resultados en nuestro modelo, que pasan por cómo presentamos nuestras variables al algoritmo¹⁵:

- Codificación: ciertos algoritmos, por la manera en la que están contruidos solo aceptan la entrada de datos bajo una forma estrictamente numérica, como el caso de las redes neuronales. Ello implica la transformación de información originalmente categórica en numérica.

Ejemplo:

Una variable categórica como color puede codificarse como un vector numérico. Si la variable color está compuesta por: rojo, azul, verde; podemos transformar esta información en un vector dicotómico en el que el primer, segundo y tercer elemento correspondan a si la observación presenta dicho color.

Así cada observación quedaría codificada de tal manera:

$$rojo = \vec{v}(1,0,0) ; azul = \vec{v}(0,1,0) ; verde = \vec{v}(0,0,1)$$

El mismo caso sucede en otro tipo de algoritmos en los que tan solo se permite información categórica.

- Intervalos (*bucketing*): cuando disponemos de datos numéricos, es posible que debido al amplio rango de la variable estudiada no aporte gran información e interese darle un tratamiento como categórica, permitiendo además dinamizar el aprendizaje del algoritmo al precisar menos computación. Para ello agrupamos según intervalos, tratando a la variable como ordinal.

Ejemplo:

Una variable numérica continua como la renta pueda agruparse por deciles, o por clase baja, media y alta, para observar el efecto de ciertos niveles de renta.

- Estandarización y normalización: ambas técnicas transforman el rango de valores para intervalos más fácilmente manipulables. En la estandarización vimos que se sustituye la distribución estadística original de la variable por un normal, sin embargo, en la

¹⁵ *Ibíd.* ps. 54-58

normalización esto no es así. Simplemente acota un rango de valores para que queden definidos por intervalos [0,1] o [-1,1].

Ejemplo:

Estandarización:

$$\bar{x}_i = \frac{x_i - \mu}{\sigma}$$

Normalización:

$$\bar{x}_i = \frac{x_i - \text{mín}}{\text{máx.} - \text{mín.}}$$

- NAs: otro gran problema en la práctica de la modelización son los valores omitidos.

Existen tres soluciones:

- Eliminar las observaciones con valores que falten. Supone menor dificultad en la manipulación, pero puede perderse información relevante de no contar con un dataset lo suficientemente amplio.
- Usar algoritmos que permitan trabajar con valores omitidos.
- Usar técnicas de imputación de datos, como aproximar el posible valor que debería mostrar el valor omitido. Por ejemplo, en datasets de series temporales sustituyendo por la media aritmética o utilizando el último valor conocido de la serie.

III. Entrenamiento de nuestros datos

La habilidad del machine learning es extraer información de nuestros datos para identificar patrones y/o relaciones. Para ello es necesario aportar un conjunto de datos, nuestro dataset de entrenamiento, con el que por cada nueva observación el aprendizaje automático ajusta su modelo. Dependiendo del tipo de aprendizaje que empleemos existen cuatro tipos distintos de modelos de machine learning¹⁶ (para mayor claridad recomiendo visitar la Tabla 1 de la p. 13):

- Aprendizaje no supervisado: se interpreta y analiza únicamente los inputs para crear nuevos outputs que etiquetan cada observación, teniendo esta clasificación como criterio que compartan características entre sí. Conformarán conjuntos homogéneos y diferenciados a su vez de otros conjuntos, o transformaciones del vector introducido

¹⁶ *Ibíd.* p. 4

como input. Los outputs nos los dará el propio modelo y serán etiquetas informando de a que clúster en concreto pertenecerá cada una de las observaciones. Existen diferentes tipos¹⁷:

1. **Estático**, dentro del cual encontramos:
 - 1.1 Particional: divide nuestro dataset en conjuntos minimizando las distancias intracluster de las observaciones y maximizando la intercluster. Aquí los clústeres son exhaustivos y no presentan intersecciones entre sí.
 - 1.2 Jerárquico: divide en clústeres que se van agrupando por nivel de proximidad, semejanza entre clústeres.
 - 1.3 Otros: como *el model-based* o *grid-based* o de densidad
2. **Series temporales**, implica ciertas complicaciones en su tratamiento estadístico debido a la temporalidad, normalmente se utiliza para su mejor manejo técnicas de reducción de dimensionalidad, entre las que diferenciamos:
 - 2.1 Crisp clustering: parten nítidamente nuestras observaciones. Son a su vez:
 - 2.1.1 Jerárquico
 - 2.1.2 Particional
 - 2.2 Fuzzy clustering: asigna grados de pertenencia a los clústeres. Si en el *crisp clustering* una observación solo puede vincularse en exclusiva a un clúster, en *fuzzy clustering* puede pertenecer en distinta medida a varios clústeres.

Este aprendizaje es habitualmente usado para segmentar, por ejemplo, identificando grupos de clientes a los que dirigir una campaña de marketing. Con esta técnica agruparemos elementos de nuestra muestra que compartan semejanzas y se diferencien de otras agrupaciones. Para la elaboración del modelo solo se utilizan los inputs. Este tipo de machine learning es comúnmente desarrollado por algoritmos *clustering* de K-medias.

- Aprendizaje supervisado: utiliza los inputs, pero también los outputs de nuestra muestra para elaborar un modelo predictivo. Los outputs serán el valor concreto de la variable a predecir por nuestro modelo. De esta técnica derivan dos tipos de modelos:
 1. de **clasificación** cuando la variable a predecir es categórica y el numero de clases de la variable es finito.
 2. o de **regresión** cuando es numérica.

¹⁷ (Sardá-Espinosa, 2018)

En el aprendizaje supervisado el algoritmo concreto usado para modelizar es escogido por la máquina por ser el que mejor relaciona cada input con su output, pero de entre una familia de algoritmos escogida a priori el modelizador. La relación se refleja en una expresión matemática de entre una de esas familias de funciones, algunas de estas familias son: redes neuronales; árboles, bosques o junglas de decisión; funciones bayesianas, regresiones lineales, etc.

- Aprendizaje semisupervisado: nuestro dataset contiene observaciones en las que existen pares de inputs y outputs, y otras en las que tan solo figuran inputs. El objetivo es deducir el output que correspondería al respectivo output, pero en el proceso se utilizaría también la relación entre inputs desclasificados (inputs sin su respectivo output). En el aprendizaje semisupervisado aun así también utilizamos las observaciones con inputs sin labels (sin outputs), etiquetas para establecer relaciones, ya que puede ser de utilidad igualmente para relacionar inputs entre sí.
- Aprendizaje reforzado: es utilizado en la resolución de problemas secuenciales en los que cada decisión tomada por el algoritmo condiciona la siguiente. En el aprendizaje reforzado el algoritmo idóneo es aquel que optimiza la recompensa media esperada, buscando así la mejor política de acciones.

Es una técnica todavía en ciernes, pero prometedora, con un potencial mayor que ninguna otra, pero actualmente los éxitos en este campo han sido escasos y sometidos a restricciones muy estrictas y delimitadas, sin ir más allá de entornos cerrados como videojuegos.

Los modelos que centrarán este trabajo son el supervisado y no supervisado, por acotar los límites del estudio y dado que son los más desarrollados hasta ahora.

Entre ellos, la diferencia radical es que en el aprendizaje supervisado no se utilizan todos los datos para construir los modelos, solo se utiliza un porcentaje de ellos (normalmente se usan la mayoría de los datos para construirlo, entorno al 80%) para construir el modelo, entrenando los algoritmos.

Mientras que el resto de los datos no utilizados a priori, el 20%, se usan a posteriori para evaluar la capacidad predictiva del modelo, en base a diferentes métricas y métodos estadísticos. Con ellos medimos como de bueno es el modelo construido. Estos son el conjunto de validación.

IV. Evaluación del modelo

El paso final examina la capacidad predictiva de nuestro modelo, paso este que se excluye en la metodología de la elaboración de clústeres al no ser modelos supervisados. Clusterizar implica clasificar las observaciones, dentro de una nueva variable categórica creada. Esta variable recoge los clústeres a los que puede pertenecer. Cada observación se asigna a una etiqueta, un clúster. Pese a todo, existen los CVI (*Cluster Validity Indexes*) que miden como de buena es la partición, según distintos criterios, pero no existe un criterio objetivo.

La corrección de nuestro modelo se realiza comparando los outputs reales con los outputs predichos por nuestro modelo. Recordemos que los inputs y outputs utilizados para evaluar la bondad de nuestro modelo vienen de los datos muestrales no utilizados para modelar. Dependiendo de si nuestro modelo es de regresión o clasificación compararemos distintas métricas para seleccionar nuestro modelo definitivo, de entre los modelos elaborados mediante las distintas familias de funciones. Por simplicidad nos quedaremos con tan solo algunas de estas medidas auxiliares:

Métricas para selección de nuestro modelo, dependiendo del tipo de modelo, si es de:

- Clasificación
 - AUC (*Area Under the Curve*): en modelos binarios su valor es la integral de la curva ROC, curva que relaciona la capacidad clasificatoria de un modelo binario, comparando las ratios de falsos positivos con los de verdaderos positivos, de los outputs predichos. AUC toma valores de entre 0 y 1, y cuanto más cercano a 1 mejores predicciones hará el modelo.
 - Matriz de confusión: usada principalmente en modelos de clasificación, en los que una matriz de dimensiones 2 x 2, por ejemplo, en un modelo binario, compara el error en la predicción de 1 o de 0, o del total de casos. Esta comparativa nos interesará, por ejemplo, cuando busquemos modelos que tengan mejor capacidad predictiva de un posible output concreto sobre otro¹⁸. Es decir que prediga mejor un posible caso que otro. El modelo elegido será el de menor error, el que más acierte acorde a el interés de nuestro análisis.
- Regresión
 - RMSE (*Root Mean Square Error*): medida de la raíz cuadrada del error medio, o sea la media aritmética de nuestra muestra de la diferencia entre el valor real de

¹⁸ En modelos para predecir si alguien padece una enfermedad o no va a pagar un crédito, nos interesa un modelo que sea mejor para predecir si la padece o si no va a haber impago que al revés.

cada output y su respectivo valor predicho, al cuadrado. Estadísticamente se eleva al cuadrado y luego se aplica la raíz cuadrada para facilitar su tratamiento matemático, que a priori debería realizarse con el valor absoluto del error, ya que algunas de esas diferencias entre el valor real y el predicho pueden ser negativas. Nos quedaremos con el modelo que presente menor error de media sobre el conjunto muestral, esto es con el RMSE más próximo a 0.

En cualquier caso, resulta conveniente explicar que la labor del investigador aquí vuelve a jugar un papel importante, ya que dependiendo del propósito que persiga puede interesarle más desarrollar métricas personalizadas, *ad hoc*, para el modelo que esté desarrollando.

V. Otras consideraciones

Existen una serie de aspectos a tener en cuenta una vez hemos construido nuestro modelo final. Y es que es posible que pese a que tengamos un modelo que sea capaz de predecir con gran precisión los outputs deseados no sea un modelo operativo ¿Por qué?

i. Sobreajuste:

Aunque el modelo funcione muy bien para la muestra empleada, para nuestro dataset, podría suceder que, sin embargo, a la hora de extrapolarlo a otros conjuntos poblacionales yerre. Esto suele ser debido a que las variables que funcionan para nuestro conjunto de entrenamiento no son realmente relevantes para esos otros datos.

La consecuencia principal es que nuestro modelo solo es útil para nuestra muestra, pero no podemos extraer conclusiones para nueva información.

Para solucionarlo podemos optar por

- reducir el número de variables que incluimos en nuestro modelo final;
- emplear modelos más sencillos, comúnmente se usan los modelos de regresión lineal porque son los que menos sobreajuste sufren, los más simples;
- regularizar, término que engloba determinadas técnicas para construir modelos menos complejos, por ejemplo, penalizando algoritmos más complejos que otros que incluyan menos variables.

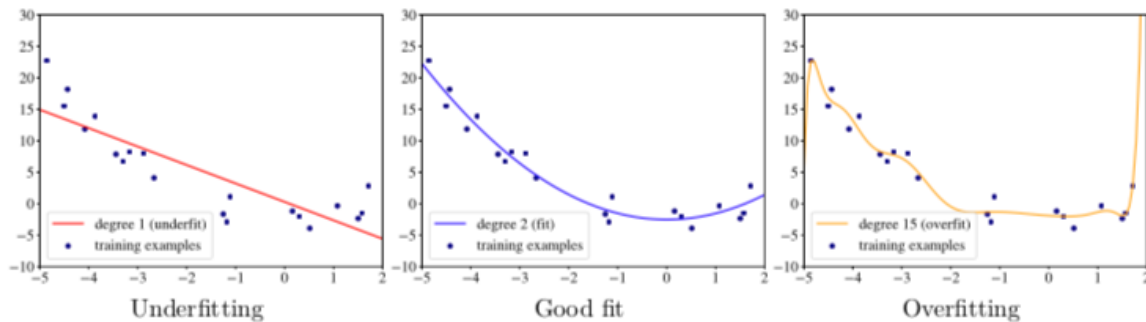
El sobreajuste supone la mayor amenaza para el machine learning, ya que los algoritmos tienden al sobreajuste. Especialmente cuando en el dataset encontramos muchas variables, pero un número reducido de observaciones

ii. Infraajuste:

Por el contrario, es posible que el modelo yerre para nuestro conjunto de entrenamiento, ello se debe principalmente a dos razones:

- Las variables incluidas no aportan información, no existe una relación con el output.
- La información muestral sea escasa, por lo que puede ser recomendable usar modelos más complejos.

Gráfico 3 Infraajuste y Sobreajuste (Fuente: "the hundred-page machine learning book", p.62, Figure 2)



VI. Selección del Modelo

Elaborados los distintos modelos entrenados el mejor se elige en base a las métricas, y de ellos serviremos para elaborar nuestras predicciones sobre el futuro en base a los datos históricos, por ello hemos de tener en cuenta que la validez del modelo solo rige mientras se cumplan las mismas condiciones en nuestros datos y su distribución que se consideraron al entrenarse. En otras palabras, existe una limitación adicional a la posibilidad de predecir, han de cumplirse las condiciones originales a su extracción.

e. Aplicaciones

Las aplicaciones del machine learning están centradas como hemos visto en la elaboración de modelos automatizados, es decir que la principal ventaja que presenta el machine learning es la posibilidad, gracias a los avances tecnológicos en la mejora capacidad de computación y procesamiento de nuestros ordenadores, de crear modelos óptimos y mecánicos con los que entender la realidad, en base a los datos recogidos.

En otras palabras, simplifica todo el proceso de modelización al ser el propio ordenador quien se encarga de escoger el algoritmo que más se adecúe a nuestros datos, dejando a la labor del investigador su mera interpretación.

Modelizar, como fue adelantado, trata de reconciliar las interacciones entre distintas variables, en otras palabras: cómo dependen unos fenómenos de otros en nuestro mundo exterior y la dinámica que rige entre ellas; al no poder ser capaces de entenderla con exactitud matemática.

Las aplicaciones de esta técnica son aún recientes y novedosas, y su funcionalidad como herramienta de entendimiento depende del tipo de modelo construido. Retomando la división inicial de los distintos tipos de modelos que podemos construir con el machine learning, distinguimos su utilidad según el tipo de aprendizaje y el objetivo de la investigación:

- **Clustering**: divide nuestra población en conjuntos de características similares. Esto nos permite hacer un análisis diferenciado sobre estas características propias y diferenciadas de cada grupo, incluso dividir en más subgrupos para investigar individualmente cada subconjunto. Su utilidad es precisamente esa, que nos permite diferenciar, en agrupaciones, una población.

El clustering es muy utilizado en marketing para segmentar mercados, distinguir tipos de consumidores y recomendar productos. Si por ejemplo una tienda dispusiese de la suficiente información sobre sus clientes podría elaborar varios clústeres en los que distinguiese las preferencias en las compras de cada grupo, permitiéndole ajustar su oferta para que cada cliente reciba productos personalizados según las preferencias del grupo al que pertenezca.

- **Clasificación**: predice variables categóricas de manera que podemos realizar predicciones fundamentadas por el modelo y con cierto grado de certidumbre sobre un parámetro categórico de una nueva observación, en base a los nuevos inputs de la nueva observación. Siempre que la información muestral sea extrapolable a la población.

Un modelo de clasificación permite conocer si un nuevo paciente es probable que tenga o no diabetes, en base a sus datos médicos. Los datos médicos históricos de otros pacientes con diabetes, los inputs, pueden ayudar a predecir si un nuevo paciente también es diabético, el output. La variable categórica en este caso sería dicotómica nominal, por ejemplo, asociando 1 a tener diabetes y 0 a no tener. El modelo nos dice a qué tipo pertenece la nueva observación.

- **Regresión**: predice variables numéricas. El modelo de regresión predice un valor concreto, o un rango de valores, de la variable endógena, sujeto/s o no a un margen de error. El error mide la diferencia entre el valor real y el predicho.

El uso de la regresión nos permite aproximar a un valor concreto el output de la variable. Algunos ejemplos del uso de modelos de regresión incluirían: tratar de predecir el valor futuro de una acción en base a una serie de factores como, los movimientos de mercados relacionados, la tasa de crecimiento de la industria, o el precio de un inmueble dadas sus características de mercado.

Hemos de tener en mente que todos los modelos son falsos por definición. Esto acarrea una consecuencia principal y es que, a la hora de aplicar un modelo, por muy ajustado a la realidad que sea siempre **es preciso cuantificar el error**, particularmente en modelos predictivos. Aunque sea de manera aproximada, para dotar de mayor seguridad y anticipar una segura toma de decisiones en base a ellos. Si el modelo no es lo suficientemente bueno tendremos que descartarlo.

También hay que establecer bajo qué condiciones es operativo el modelo, de otro modo no podemos aplicarlo bajo circunstancias sustancialmente diferentes. La solución a este problema pasa por un análisis estadístico de nuestro set de entrenamiento.

En el caso de que elaborásemos un modelo predictivo sobre el valor de la acción de una empresa en un mercado con una volatilidad histórica determinada según un intervalo reducido, es decir con una volatilidad baja, no sería posible utilizarlo ante un cambio disruptivo en la dinámica del mercado, que implique un incremento sustancial de la volatilidad de forma que haga inoperativo nuestro modelo por no recoger el impacto de ese incremento. Perdería su capacidad predictiva.

Del mismo modo que si intentáramos predecir el movimiento de precios de una acción en un contexto de recesión económica elaborando un modelo con datos recogidos durante un periodo de expansión. Las circunstancias no son extrapolables.

3. Trading

“[...] El ‘trading’ consiste en la compraventa de activos cotizados con mucha liquidez de mercado (sobre todo, acciones, divisas y futuros) en un mercado financiero electrónico y regulado. Su objetivo es obtener un beneficio económico cuando la operación genera una plusvalía.”¹⁹

El trading es el conjunto de operaciones a corto plazo utilizado por los inversores para generar un beneficio mediante transacciones financieras. Sus elementos característicos son.

- **Compraventa:** implica bien la compra y/o la venta de un producto en un mercado regulado. Esta puede ser inmediata o mediata dependiendo del instrumento de inversión utilizado: contratos de futuros, opciones, índices u otros derivados, divisas, o con activos o pasivos (bonos) mediante la toma de posiciones alcistas o bajistas. Esto son posiciones largas o cortas, las posiciones mas simples que se pueden tomar, con las que el trader construirá su estrategia de inversión. Se pondrá largo en un activo o derivado si espera que se revalorice o corto si espera que este se deprecie.

En todo caso la compraventa tiene lugar en mercados regulados.

- **Mercados:** los mercados de trading son mercados financieros, porque sus productos están sujetos a un rendimiento económico temporal o a un tipo de interés. Los mercados financieros están compuestos por dos tipos de mercados:
 - a. Primarios, regulan la emisión de nuevas acciones y deuda publica
 - b. Secundarios, que son mercados a su vez de,
 - i. Renta variable: Valores
 - ii. Renta fija
 - iii. Derivados: Futuros

El hecho de que sean, por otro lado, regulados implica la existencia de una serie de garantías para los operadores de mercado, mediante dicha regulación por instituciones de carácter público como son la CNMV en España o la SEC en EE. UU, se limita la capacidad dispositiva de las partes. Entre otras cosas se prohíbe la información privilegiada, se elimina el riesgo de insolvencia crediticia mediante la constitución de garantías y se estandarizan los contratos.

¹⁹ [BBVA ¿Qué es el 'trading'?](https://www.bbva.com/es/que-es-trading-que-hace-falta-para-operar/) Visto el 14/06/2020 en: <https://www.bbva.com/es/que-es-trading-que-hace-falta-para-operar/>

- Corto plazo: significa que las operaciones de trading no buscan un beneficio intrínseco por la mera tenencia de las posiciones abiertas, si no que el resultado se busca cerrándolas. En otras palabras, el trading tiene como objetivo exclusivo generar una plusvalía en la compra y posterior venta, o viceversa, no busca un beneficio por la sola compra, si no mediante el intercambio.

Ello implica que el trading solo se centra prioritariamente por el valor de mercado, por tanto, en el análisis técnico de precios y no tanto el valor intrínseco, que implicaría análisis fundamental.

- Beneficio: el trading tiene una vocación eminentemente lucrativa, cada operador o trader busca maximizar sus beneficios con cada operación. Beneficios que se miden única y exclusivamente por el valor económico generado, no por ejemplo en el valor de los derechos políticos de una acción sobre la sociedad o su valor para negociar.
- Liquidez: el trading se desvincula de la inversión tradicional en la rotación de capital²⁰, la cantidad de veces que se abre una posición con el capital. Dada la gran cantidad de volúmenes de compra y venta, la ejecución de las órdenes en el trading suele ser casi inmediata lo que propicia que el capital invertido se utilice en un periodo muchas más transacciones. En inversiones largo placistas la proporción de transacciones es mucho menor.

La razón de buscar mercados muy líquidos es para maximizar las opciones del trader para cerrar o abrir posiciones, cumpliendo con el corto plazo.

a. Principios básicos

En base a la conjugación de *longs* y *shorts*, posiciones largas y cortas respectivamente, el trader puede crear posiciones sintéticas, como replicar el comportamiento de los activos subyacentes; apalancarse; cubrir los riesgos de sus posiciones mediante hedging; o adoptar posiciones *bullish*, *bearish*, *strangle*, *straddle* o *butterfly*, entre otras, si usa opciones.

Las estrategias de trading que más nos interesan a fin de observar donde puede tener cabida el machine learning son la mayoría y de hecho, pueden y suelen tener carácter concurrente. Comúnmente se utiliza²¹:

- Arbitraje: consiste en la obtención de ganancias sin riesgo de pérdidas, aprovechando las ineficiencias del mercado. En una operación de arbitraje el trader tomará una

²⁰ (Ajram, 2020) p. 16

²¹ (Staff of the Division of Trading and Markets U.S. Securities and Exchange, 2014) p. 4

posición y la contraria en un mismo activo subyacente simultáneamente, sacando un rédito económico.

- *Market making* pasivo: simultáneamente el actor toma posiciones contrarias en un mismo activo subyacente, sin embargo, el *market maker* sí que corre riesgos con su operación. Comúnmente hablaríamos de una situación de reventa en la que el beneficio no está asegurado.
- Trading estructural: explota las ineficiencias en el mercado o en otros operadores del mercado. En el caso del *flash trading* se explota la menor latencia para obtener información antes que la competencia, más tarde se explicará pormenorizadamente el *flash trading*.
- Trading direccional: toma de posiciones según el movimiento que se prevea en el valor. La estrategia eminente que será posteriormente desarrollada en la parte 4 de este trabajo.

b. Aplicaciones de estadísticas a la inversión

La utilización de modelos estadísticos en la economía es extensa dada la incapacidad de reflejar con exactitud en una fórmula matemática los comportamientos del mercado. Ya sea para datos macro o micro, la econometría nace como instrumento de aproximación en el entorno económico utilizando información histórica para elaborar perspectivas de futuro.

Dentro del campo de las finanzas la utilización de las matemáticas, la estadística y las ciencias actuariales no es ninguna novedad, ni mucho menos en los mercados financieros. Podemos encontrar sectores en los que la utilización de modelos es recurrente como en: los seguros, para fijar la prima; o el sector bancario, para clasificar el riesgo de insolvencia. En los mercados de derivados y acciones, también encontramos modelos estadísticos de valoración como por ejemplo el modelo Black-Scholes de valoración de opciones, o la Teoría de optimización de carteras de Markowitz.

En los mercados financieros, de hecho, la utilización de modelos de valoración y de algoritmos de toma de decisiones automatizadas es muy frecuente, especialmente en los fondos cuánticos con el denominado *flash trading* o el *High Frequency Trading*²² (HFT), que según la SEC es uno de los mayores *game changers* recientes de la estructura de mercado, y que según sus estimaciones involucra la mitad del total de acciones cotizadas en EE. UU.²³

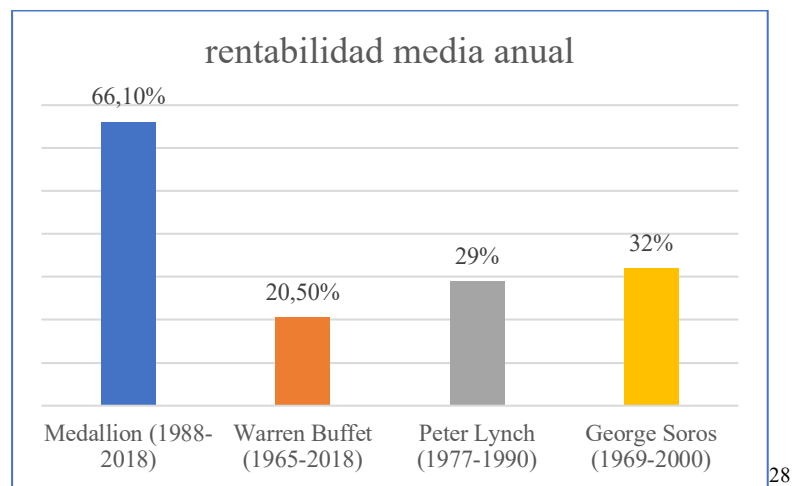
²² (Lacalle, 2013) p. 60

²³ (Staff of the Division of Trading and Markets U.S. Securities and Exchange, 2014) p. 4.

Como ejemplo de este proceso de algoritmización de la economía e implementación de la IA, fondos como AIEQ²⁴, un ETF²⁵ han logrado doblar la rentabilidad de mercado del índice tecnológico estadounidense S&P500 (un 13% frente al 7%, a YTD²⁶), de la mano de la compañía EquBot.

Pero sin duda alguna, el ejemplo más exitoso de ello es Renaissance Technologies mediante su fondo Medallion que obtuvo un 66% de rentabilidad media anual desde 1988²⁷. Dicho fondo consiguió batir las rentabilidades anuales medias de fondos afamados por batir anualmente la rentabilidad del mercado como Berkshire Hathaway o Magallan. Todo ello precisamente con la pionera utilización del machine learning y la algoritmización.

Gráfico 4 Comparativa de los mejores inversores del mundo (Fuente: elaboración propia a partir de los datos de El País)



I. Análisis de valores

Pero, antes de ver cómo puede tener cabida la estadística en la inversión, conviene recordar las bases sobre las que se erigen cualquier estrategia de inversión: el análisis fundamental y el técnico.

²⁴ [AI-Powered Fund Thrashing S&P 500 Is Wary of Cyclical Bounce](https://www.bloomberg.com/news/articles/2020-10-21/market-beating-robot-fund-manager-shys-away-from-cyclical-bets). Visto el 25/10/2020 en: <https://www.bloomberg.com/news/articles/2020-10-21/market-beating-robot-fund-manager-shys-away-from-cyclical-bets>

²⁵ *Exchange Traded Fund* (ETF), fondos de acciones cotizadas que a su vez cotizan en los mercados.

²⁶ *Year To Date* (YTD), un año medido desde la fecha de referencia.

²⁷ [El fondo que convierte 100 dólares en 400 millones en 30 años](https://cincodias.elpais.com/cincodias/2019/11/15/mercados/1573831472_956898.html). Visto el 21/09/2020 en: https://cincodias.elpais.com/cincodias/2019/11/15/mercados/1573831472_956898.html

²⁸ La rentabilidad de Medallion está medida antes de las comisiones a cargo de los brókeres, muy altas, dada la enorme rotación de capital y la gran cantidad de transacciones. Aun así, supera al resto de fondos.

- **Análisis fundamental:** comúnmente realizado usando modelos no necesariamente estadísticos como el de Discounted Cash Flows, que considera la capacidad de una empresa de: generación futura de beneficios, solvencia y crecimiento. También se utilizan otros modelos la valoración por múltiplos, el modelo de Graham, el CAPM (*Capital Asset Pricing Model*) o el modelo de Gordon-Shapiro. En base a estos criterios el análisis fundamental fija un valor intrínseco del valor, que necesariamente no tiene por qué coincidir con su valor de mercado.

Esta divergencia es explotada por el inversor para tomar posiciones cortas o largas partiendo de la premisa que se producirá una corrección, en la que el valor de mercado convergirá con el valor intrínseco²⁹.

Se orienta más al largo plazo. Este método ha sido el seguido por los grandes inversores tradicionales como Warren Buffet o Peter Lynch.

- **Análisis técnico:** solo utiliza la información OHLC³⁰ histórica del valor en el mercado para determinar si un valor está barato o no. Con el análisis técnico se trata de buscar tendencias, resistencias, rebotes o patrones en la fluctuación de precios. Se orienta al corto plazo. Este es principalmente el análisis utilizado por Medallion.

Ambos análisis suelen ser utilizados por los inversores y en ambos pueden ser de utilidad los modelos estadísticos, sin embargo, este trabajo se va a centrar en exclusiva en el desarrollo del machine learning en el campo del análisis técnico dada la complejidad técnica de establecer una relación matemática, al carecer de certeza sobre las variables explicativas y su importancia. No sabemos exactamente qué información es relevante. Recordemos que esta es la antes mencionada premisa por la que empezar a considerar la utilidad de cualquier modelo estadístico. Porque la estadística es la mejor aproximación posible que podemos hacer para medir el error en nuestras decisiones.

El análisis técnico trata entonces de anticipar las fluctuaciones de precios en base a la estructura interna del propio mercado tomando como referencia los movimientos de precios históricos producidos en situaciones estructuralmente semejantes, en la actividad corriente del mercado.

El propósito es predecir la tendencia que seguirán los precios.

Sin embargo, la habilidad predictiva del análisis técnico requiere de una condición necesaria, anteriormente mencionada, para extrapolar los resultados obtenidos con nuestro set de datos de entrenamiento. Y es que, para considerar que nos encontramos ante situaciones

²⁹ (Rankia, 2016)

³⁰ En un valor los precios de apertura, cierre, y el máximo y mínimo, así como el volumen de la sesión.

estructuralmente semejantes y en la actividad corriente del mercado, la distribución estadística de nuestros datos de entrenamiento tiene que ser igual o semejante a la que sigan los nuevos inputs que introduzcamos en nuestro modelo ya construido.

II. La teoría del Dow

Ya en 1993 se hacía referencia al chartismo, la utilización de métodos gráficos, como el “principal instrumento de análisis técnico”³¹, e incluso hoy algunas estrategias de trading se fundamentan en él. Desarrollada por el chartismo, la teoría del Dow busca conjuntos de valores que se muevan en paralelo. Si conocemos la tendencia del conjunto conoceremos la tendencia de los valores que lo conforman pudiendo obtener un rédito. Esta teoría parte de una serie de principios:

- Las medias descuentan todo lo que puede ser conocido.
- Existen tres tendencias del mercado: primaria, secundaria y diaria.
- Las tendencias principales tienen tres fases: acumulación, intermedia y distribución.
- Las medias deben confirmarse entre sí.
- El volumen debe confirmar la tendencia.
- Una tendencia perdura hasta ofrecer signos definitivos de haber cambiado.

La teoría del Dow ha sido ampliamente desarrollada mediante diferentes métodos como gráficos de barras, de puntos y signos, las medias móviles, el análisis cíclico, etc.

Partiendo del rasgo principal de esta teoría: la construcción de conjuntos de valores que sigan tendencias paralelas; el machine learning, en concreto por medio de la clusterización, parece la técnica perfecta para actualizar la teoría del Dow a las capacidades tecnológicas de nuestros días. Precisamente esta es la característica principal del clustering, etiquetar las observaciones según el clúster al que pertenezcan. Sin duda el Big Data y el análisis de datos pueden ayudar a elaborar perfiles de valores que compartan propiedades comunes, pudiendo por ejemplo elaborar conjuntos con los retornos que ofrezcan cada valor como criterio diferenciador.

III. Análisis técnico-estadístico

En este tipo de análisis se trata de mecanizar las órdenes de compra o venta, relegando a un segundo plano la labor del inversor. La labor interpretativa en el análisis chartista y la teoría del Dow tiene mayor desarrollo que en el análisis estadístico, en la que los resultados numéricos amparan las decisiones finales en señales cuantificadas. En el desarrollo estadístico tienen una gran importancia los osciladores de precios.

³¹ (Costa Ran y Font Vilalta, 1993) ps. 80, 81

Estos son indicadores de la tendencia, pero tienen un carácter secundario y subordinado al análisis básico de la tendencia que siga el valor observado. La interpretación de los osciladores es que cuando un valor se acerca a sus límites extremos, superiores o inferiores, se produce un efecto rebote, en la dirección opuesta a la tendencia seguida por el valor, para compensar el exceso de compra o bien de venta.

Algunos tipos de osciladores clásicamente utilizados para determinar si existe sobrecompra o sobreventa son:

- **Relative Strength Index (RSI)**: mide la fuerza de un precio con relación a su variación al cierre.

$$RSI = 100 - \frac{100}{1+RS}, \text{ siendo } RS = \frac{Si/N}{Sr/N}$$

Si = suma del incremento de precio al cierre

Sr = suma del descenso del precio

N = Número del periodo a actuar

Si $RSI > 70$ existe sobrecompra, el precio bajará por lo que deberíamos vender.

Si $RSI < 30$ existe sobreventa, deberíamos comprar porque el precio subirá.

- **K%D estocástico**: recoge la propensión de un valor a seguir la tendencia de su precio al cierre.

$$\%K = \frac{(C - LN)}{HN - LN'}$$

Si $\%K > 80$ sobrecompra

Si $\%K < 20$ sobreventa

C = precio de cierre

LN = precio mínimo del periodo estudiado

HN = precio máximo del periodo

- **Momentos**: compara la velocidad a la que se aprecia o deprecia un valor. Algunos indicadores clásicos son el *Rate of Change* (ROC) o el *Commodity Channel Index* (CCI).
- **Indicador Moving Average Convergence Divergence (MACD)**: relaciona la diferencia entre medias móviles exponenciales de 9 y 18 sesiones, siendo una de ellas más rápida que la otra.

Vamos a ver un ejemplo clásico del análisis chartista para trading. En la siguiente imagen, el Gráfico 5, se analiza la cotización usando: bandas Bollinger, el MACD, el RSI, el CCI y el

ROC (respectivamente de arriba abajo) para estudiar los movimientos en el precio del oro (XAU) medido en dólares.

Haciendo el análisis de la gráfica un trader tradicional estudiaría los distintos indicadores para asegurarse posibilidades de éxito en sus operaciones.

Por ejemplo, en Julio el valor del oro de acuerdo con las bandas de Bollinger se encuentra en una situación de resistencias, además presenta un RSI > 70 por lo que según el chartismo debería abrir posiciones cortas en oro.

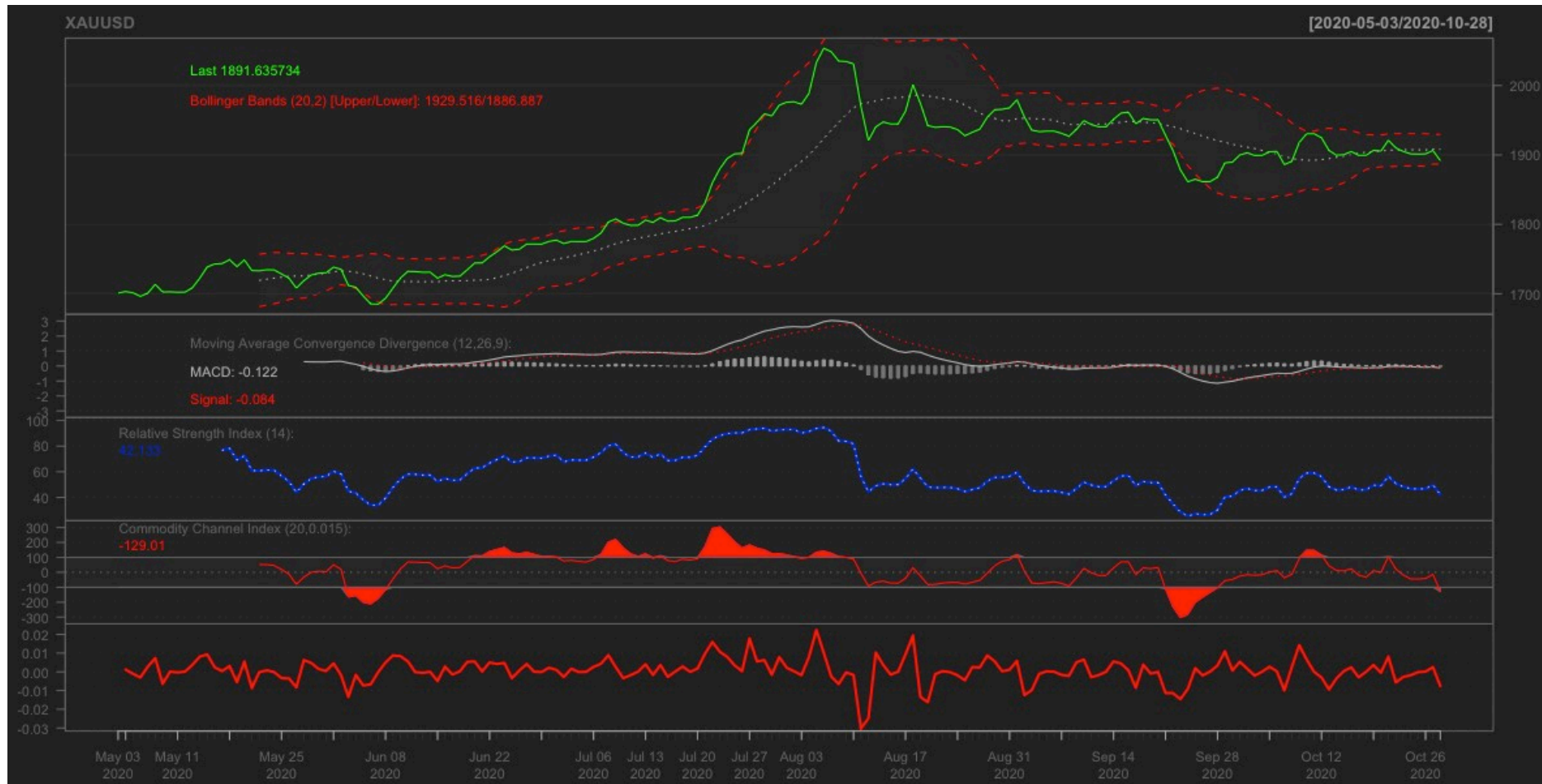
Todas estas teorías dependen principalmente de las medias móviles de los precios. Las medias comúnmente consideradas suelen ser la aritmética, o la exponencial y ponderada para otorgar mayor peso a las cotizaciones más recientes.

Suelen utilizarse las de las últimas 200, 70 y 6 sesiones simultáneamente, para determinar el comportamiento de la tendencia a largo, medio y corto plazo. En esta teoría se amparan populares estrategias de inversión como el cruce de medias³², en la que se pronostica una posible subida o bajada de la cotización dependiendo de la tendencia de las medias cuando se produce una intersección gráfica entre distintas medias móviles, siempre que las medias móviles confirmen la tendencia y dentro de unos márgenes determinados.

Como conclusión, la estadística amparada por los avances computacionales ha sido muy empleada en la inversión. Pero todos los análisis expuestos revisten un carácter más bien complementario que autónomo, siendo en la práctica compatibles. De hecho, es más que aconsejable la realización simultánea de análisis técnicos y fundamentales en combinación.

³² (Ajram, 2020) ps. 61 y 62

Gráfico 5 Chartismo aplicado al oro en USD (\$) (Fuente: elaboración propia en R)



IV. La nueva revolución tecnológica

El machine learning, el Big Data, y los nuevos tratamientos de datos, relacionados con la estadística son el pasado, el presente y el futuro de la inversión. De hecho, son ya una realidad disruptiva de este sector de la economía global y suponen en casos como el de Medallion una ventaja competitiva capaz de batir a cualquier inversor humano.

Reticentemente el trading tradicional, obsoleto, trata de desajustar los modelos de trading algorítmico con técnicas en el borde de la ilegalidad como, entre ellas, el *spoofing*: lanzando falsas ofertas en los mercados para descuadrar los algoritmos del *high frequency trading*, que se verían forzados a recalcular la oferta y demanda del mercado; y posteriormente cancelando las ofertas³³.

Sin embargo, las autoridades de la competencia persiguen el falseamiento del mercado. La decadencia del trading humano es ya un hecho a gran escala, pudiendo el algoritmo en cuestión de segundos igualar y cerrar posiciones, es decir puede replicar traders humanos batiéndoles en velocidad de síntesis de los nuevos datos y de emisión de órdenes.

Podemos afirmar que estamos ya en el ocaso de la inversión manual, que se está viendo relegada los procesos de inversión automatizada, con el trading algorítmico o el trading computacionalmente asistido.

c. Trading algorítmico

Visto el desarrollo del trading tradicional y sus fundamentos vamos a revisar la actualidad del trading algorítmico. Sabemos que el trading algorítmico no es más que la programación de una serie de reglas según la que ejecutar o no distintas órdenes.

La definición de la novedosa directiva MiFID II³⁴ regula el trading algorítmico en base a la siguiente definición en su artículo 4.1.39:

“«negociación algorítmica»: la negociación de instrumentos financieros en la que un algoritmo informático determina automáticamente los distintos parámetros de las órdenes (si la orden va a ejecutarse o no, el momento, el precio, la cantidad, cómo va a gestionarse después de su presentación), con limitada o nula intervención humana [...]”

³³ [Inside the JPMorgan Trading Desk the U.S. Called a Crime Ring](https://www.bloomberg.com/news/articles/2020-09-28/inside-the-jpm-precious-metals-desk-called-a-crime-ring-by-prosecutors), Visto el 28/09/2020 en: <https://www.bloomberg.com/news/articles/2020-09-28/inside-the-jpm-precious-metals-desk-called-a-crime-ring-by-prosecutors>

³⁴ DIRECTIVA 2014/65/UE

Hoy por hoy los avances computacionales están revolucionando el trading algorítmico basado en los modelos estadísticos mediante dos técnicas mencionadas en las que merece la pena ahondar, ya que en ellas se arraiga el actual estado del arte del trading algorítmico, alguno de ellas en absoluto exenta de polémicas.

II. Flash trading

Dada la diferencia tecnológica entre los operadores del mercado es posible que algunos de ellos se beneficien de una mayor rapidez de visualización, y por ende disponiendo de una ventaja frente a otros al tener más tiempo de reacción.

La latencia o diferencia temporal es mínima, apenas de “fracciones de segundo”³⁵. Aun así, es suficiente para que los *market makers*³⁶ puedan beneficiarse de esta información, efectuando arbitrajes sobre los precios.

Los propios mercados primarios eran en origen quienes ofrecían a los *market makers* esta información a cambio de comisiones.

Dados los principios de libre mercado se rompe el principio de información simétrica en perjuicio del resto de operadores, por lo que las autoridades regulatorias y en concreto la SEC han desautorizado su uso mediante la norma Reg. NMS. El *flash* trading se ha considerado una amenaza para conseguir mercados eficientes en un contexto de competencia perfecta. Sin embargo, sus defensores argumentan que las diferencias tecnológicas entre inversores no van a desaparecer y que el *flash* trading proporciona liquidez y mejores precios al mercado.

Actualmente sigue existiendo vestigialmente, pero en un número reducido de mercados cotizados.

III. High Frequency trading (HFT)

Supone la principal técnica, pero no la única de desarrollo del trading algorítmico. Históricamente sea ha beneficiado del *flash* trading, ya que ambos son compatibles, sin embargo, pese a la decadencia del flash trading el HFT ha perdurado.

No es *per se* una estrategia de trading si no una manera de ejecutar estrategias de trading clásicas.

También regulado por la MiFID II, se define legalmente en la UE por las siguientes características, en el art. 4.1.40:

³⁵ *Flash Trading*, Visto el 16/10/2020 en: <https://www.investopedia.com/terms/f/flash-trading.asp>

³⁶ Operadores del mercado secundario, normalmente actúan de brókeres ofreciendo posiciones contrarias simultáneamente (*bids* y *offers* al mismo tiempo) se contraponen a los brókeres DMA (*Direct Market Access*)

- a. una infraestructura destinada a minimizar las latencias de la red y de otro tipo, que incluya al menos uno de los siguientes sistemas para la introducción de órdenes algorítmicas: localización compartida, ubicación próxima o acceso electrónico directo de alta velocidad; L 173/384 Diario Oficial de la Unión Europea 12.6.2014 ES
- b. un sistema que determina la iniciación, generación, encaminamiento o ejecución de las órdenes sin intervención humana para cada negociación u orden, y
- c. elevadas tasas de mensajes intradía que pueden ser órdenes, cotizaciones o anulaciones;

Sin embargo, esta definición no es unánime a escala global. El tratamiento dado por la SEC en EE. UU. ha ofrecido a lo largo del tiempo diversas definiciones del HFT para estudiar entre otras la malignidad o no del HFT para los mercados.

Sin llegar a una conclusión clara ofrece cinco características³⁷ habitualmente vinculadas al HFT:

1. El uso de programas extraordinariamente rápidos y sofisticados para generar, enrutar y ejecutar órdenes. Los algoritmos y los modelos en los que se sustentan
2. Utilización de los servicios de co-localización y de los datos individuales que ofrecen los intercambios y otros para minimizar la red y otras latencias³⁸.
3. Plazos muy cortos para establecer y liquidar las posiciones. Normalmente intradía.
4. Presentación de un gran volumen de órdenes que se cancelan poco después tras la sumisión.
5. Terminar el día de negociación en una posición tan plana como sea posible (es decir, que no llevan posiciones significativas, sin posibilidad de *hedging*, cubrirse tomando posiciones contrarias, durante la noche).

Como vemos ambas definiciones comparten similitudes, pero para mayor claridad cuando hablemos del HFT nos referiremos a operaciones de trading algorítmico de: grandes volúmenes de operaciones, con gran rotación de capital, a gran velocidad y baja latencia.

La importancia actual del HFT es gigante. Como vimos fondos cuánticos como el de Renaissance Technologies han conseguido desbancar, y por mucho, nuestros paradigmas de inversores exitosos. Nos encontramos con que esta estrategia es imbatible frente a la obsoleta inversión tradicional, al menos obsoleta en tanto en cuanto no consiga integrar en sus fundamentos teóricos el rápido avance tecnológico que está teniendo lugar.

³⁷ (Staff of the Division of Trading and Markets U.S. Securities and Exchange Commission, 2014) p. 4.

³⁸ Latencia referida al intervalo de tiempo transcurrido entre la actualización de la información del mercado hasta que esta es accesible por otro operador.

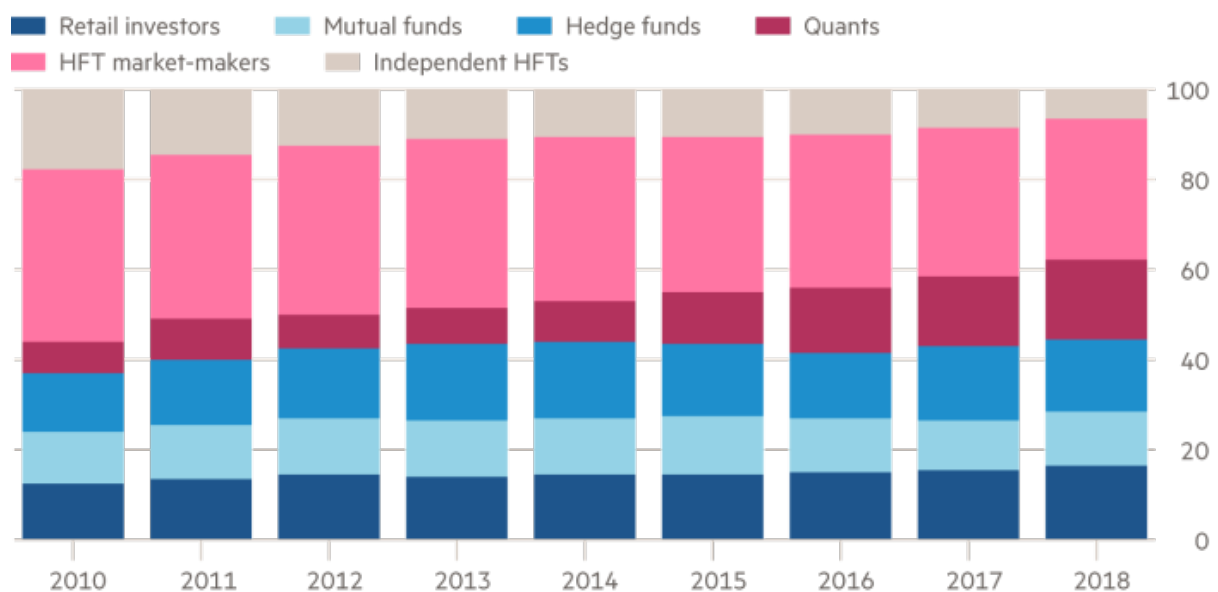
Las instituciones financieras hegemónicas tradicionales se están viendo amenazadas por este avance. Como es de esperar, viéndose en jaque su supremacía en los últimos años están proliferando una serie de acciones fraudulentas (ilegales, o alegales de dudosa ética cuanto menos), como contramedidas al trading algorítmico para salvaguardar su posición privilegiada. Antes hablaba de *spoofing*, pero existen muchas y diversas formas de manipulación de mercados (*ramping, quote stuffing, layering...*) con las que cada vez con mayor sutileza las viejas potencias intentan resistirse al cambio; por no hablar de la presión que este sector financiero tradicional ejerce sobre las instituciones políticas, judiciales, gubernamentales y estatales, en forma de *lobbying*.

Como vemos en la Gráfica 6 el peso del trading algorítmico en EE. UU. supone cerca de la mitad de todas las transacciones. Su importancia es gigantesca.

Gráfico 6 Peso de los operadores en las transacciones de acciones en EE.UU. (Fuente: Financial Times 8/1/2019)

Quants and high-frequency traders account for over half of all US equity trading

Market share (%)



Source: Tabb Group
© FT

4. Aplicación del machine learning al trading: Ejemplos prácticos

El objeto de este apartado es descubrir potenciales aplicaciones concretas del machine learning al trading. Dado que la utilidad de esta técnica reside en crear modelos automatizados, construiremos modelos *ad hoc* para tratar de explotar su potencial en el trading.

En el estudio inicial del machine learning destacamos la existencia de los dos grandes tipos de modelos más frecuentes, hoy por hoy: los supervisados y no supervisados.

a. ¿Qué modelo utilizar?

A fin de predecir los movimientos de mercado los modelos que nos servirán son los supervisados. Pero, por otro lado, ¿pueden ser los no supervisados también de utilidad? Sí, por ejemplo, desarrollando la teoría del Dow bajo una clusterización de distintos valores para distinguir conjuntos de valores que sigan tendencias paralelas, o un clustering jerárquico de valores que ordena los valores de una cesta según mayor o menor proximidad en términos de fluctuaciones, volumen, etc.; o diversificando una cartera en clústeres que sigan distintas tendencias para protegerse de la alta volatilidad.

Dentro de los supervisados debemos elegir si nuestro modelo ha de ser de regresión o clasificación. Ello dependerá de la variable que queramos predecir, de si es categórica o numérica.

Aquí surge una primera vicisitud y es que la solución a un problema resuelto por esta técnica puede hallarse por ambos tipos de modelos.

Por ejemplo:

Supongamos que un trader está pensando en abrir una posición larga en un valor para así aprovechar la oportunidad de mercado y sacar un beneficio.

En base a su modelo entrenado el trader puede tratar de predecir con un modelo supervisado el valor futuro e incluso, de seguir una distribución normal la probabilidad del intervalo de la predicción, para establecer un intervalo de confianza factible. Es decir, puede elaborar una predicción del valor puntual del tipo de cambio usando un modelo de regresión.

Sin embargo, también puede valerse de un modelo de clasificación. Imaginemos que en base a sus datos crea metadatos en una nueva variable, por ejemplo, añadiendo una variable categórica dicotómica en la que “1” significará que el valor en el futuro será mayor que el spot, o “0” si se depreciará siendo el spot mayor que el forward. Clasificaría si ese valor subirá o no en un momento futuro.

Así pues, conociendo el valor forward y pudiendo predecir el spot futuro y siendo $F(x)$ nuestro algoritmo. Si:

$F(x) = 1$, entonces: $forward0 - spot1 < 0$, en cuyo caso el trader tomará una posición larga, comprando el activo.

O si:

$F(x) = 0$, entonces: $forward0 - spot1 > 0$, el trader abrirá posiciones cortas.

Dependiendo de las preferencias del trader y la bondad del modelo es preferible bien o un modelo de clasificación o regresión. Así escogeremos:

- **Clasificación:** cuando el modelo sea un buen predictor y el margen de error sea ínfimo y poco volátil. Si nuestro modelo nos aconsejase comprar, pero el margen de beneficios fuese muy escaso, y la predicción tuviese una volatilidad alta, la probabilidad de pérdidas descompensa la toma de posiciones. Además, dependiendo de la aversión al riesgo de nuestro trader puede ser posible que no esté interesado en ganancias potenciales muy escasas, para ello nos serviremos del error de predicción en el modelo. En definitiva, un modelo de clasificación solo nos informaría de si una posición produce beneficios o pérdidas, si sube o baja³⁹.
- **Regresión:** ayuda a cuantificar el resultado potencial de la posición con intervalos estadísticos. Con él tomamos una medida aproximada del valor futuro. Con la regresión podemos aproximar un valor o valores sobre los que debería situarse el valor.

Pese a todo debe recordarse que los modelos están sujetos a determinadas limitaciones en tanto en cuanto están contruidos sobre determinadas hipótesis y basándose en información histórica, precedente. Utilizar modelos para situaciones no equivalentes o estructuralmente diferentes carece de sentido, los *insights* del modelo no son aplicables a nuestro caso.

En definitiva, el modelo a utilizar depende del propósito de nuestra investigación, de qué información concreta queremos obtener.

b. Supervisado: Clasificación

En un paper⁴⁰ exploratorio realizado por investigadores de la universidad de Stanford centran el uso del machine learning y las técnicas del HFT a un mercado relativamente novedoso, el

³⁹ Si no sube ni baja entonces, por ende, se mantiene igual. Pero ¿Si se mantiene igual nos interesa operar con él?

⁴⁰ (Madan et al., 2014)

bitcoin⁴¹. De hecho, desarrollaron un algoritmo capaz de acertar con un 98,7% de precisión el signo diario del bitcoin, si al día siguiente la criptomoneda se revalorizaría o no.

Partiendo de un dataset de 26 características, nuestras variables, la labor del equipo de investigación aplica las técnicas de machine learning anteriormente mencionadas en el apartado de metodología.

En primer lugar, selecciona aquellas que puedan aportar información, que contengan poder predictivo de nuestro output. De este primer filtro extraen las siguientes 16 variables:

Tabla 2 Preselección de la variables explicativas (Fuente: “Automated Bitcoin trading via machine learning algorithms” p. 2)

FEATURE	DEFINITION
Average Confirmation Time	Ave. time to accept transaction in block
Block Size	Average block size in MB
Cost per transaction percent	Miners revenue divided by the number of transactions
Difficulty	How difficult it is to find a new block
Estimated Transaction Volume	Total output volume without change from value
Hash Rate	Bitcoin network giga hashes per second
Market Capitalization	Number of Bitcoins in circulation * the market price
Miners Revenue	(number of BTC mined/day * market price) + transaction fees
Number of Orphaned Blocks	Number of blocks mined / day not off blockchain
Number of TXN per block	Average number of transactions per block
Number of TXN	Total number of unique Bitcoin transactions per day
Number of unique addresses	Number of unique Bitcoin addresses used per day
Total Bitcoins	Historical total Number of Bitcoins mined
TXN Fees Total	BTC value of transaction fees miners earn/day
Trade Volume	USD trade volume from the top exchanges
Transaction to trade ratio	Relationship of BTC transaction volume and USD volume

Una vez definidas y seleccionadas las variables partimos el dataset en conjuntos de entrenamiento y validación. Con estos datos elaboraran algoritmos óptimos de diversas ramas y comparan los resultados obtenidos por los diversos modelos.

Tabla 3 Selección del modelo de clasificación: Resultados (Fuente: “Ibidem” p. 3)

STATISTIC	BINOMIAL GLM	SVM	RANDOM FOREST
Sensitivity (TPR)	0.9790	0.0348	1.0
Specificity (TNR)	0.9939	0.5514	0.9392
Precision (PPV)	0.9790	0.0839	0.7762
Accuracy (ACC)	0.9879	0.2716	0.9498

⁴¹ Nacida en el 2009, ha visto aumentada su liquidez y popularidad en el mercado como refugio de las monedas fiat (respaldadas por bancos centrales o autoridades monetarias como moneda de curso legal) y su inflación. A finales de 2017 inmersa en una burbuja especulativa alcanzó su valor máximo de cotización en 16.700 €.

La métrica elegida por ellos, al ser un modelo de clasificación es la precisión (*Accuracy*), la capacidad de predecir total.

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}^{42}$$

Con esta información podemos escoger entonces aquel algoritmo de todos que presenta una mayor precisión: el Modelo Linear Generalizado binomial (*Binomial GLM*).

La conclusión a la que llegan en este experimento es que, si los datos son extrapolables y el algoritmo no sufre de sobreajuste ni de *underfitting*, podemos predecir en circunstancias equiparables con un 98,79% de acierto el precio del bitcoin al día siguiente, y como traders tomar posiciones en consecuencia.

c. No Supervisado: Clusterización

La idea que subyace en el clustering es dividir nuestros datos en conjuntos, clústeres, de miembros que entre ellos compartan tantas similitudes como sea posible. Manteniendo simultáneamente la mayor diferenciación posible entre los distintos clústeres⁴³.

Vamos a desarrollar la teoría del Dow puede ser ampliamente con esta técnica, ya que el clustering nos permite elaborar conjuntos de valores que, como en la teoría del Dow, se muevan paralelamente. Ello nos permite diferenciar, por ejemplo, valores más rentables que otros, o valores directores entre sí⁴⁴.

Tomando como referencia el artículo [Stock Clustering with Time Series Clustering in R](#)⁴⁵, en el que se clusteriza las empresas que cotizan en el S&P 500, vemos que clusterizar nos permite elaborar conjuntos cuyos movimientos de precios relativos se asemejen lo más posible.

Agrupar por como crece en el tiempo, con ello podemos distinguir valores con fluctuaciones parecidas. Estadísticamente el paquete [dtwclust](#)⁴⁶ de R utilizado, usa las técnicas DTW (*Data Time Warping*) comúnmente empleadas en datos de series temporales, para agrupar según la monotonidad de las observaciones. Agrupa en función de como se mueven.

⁴² TP y TN corresponden a verdaderos positivos y negativos, FP y FN a falsos positivos y negativos; respectivamente

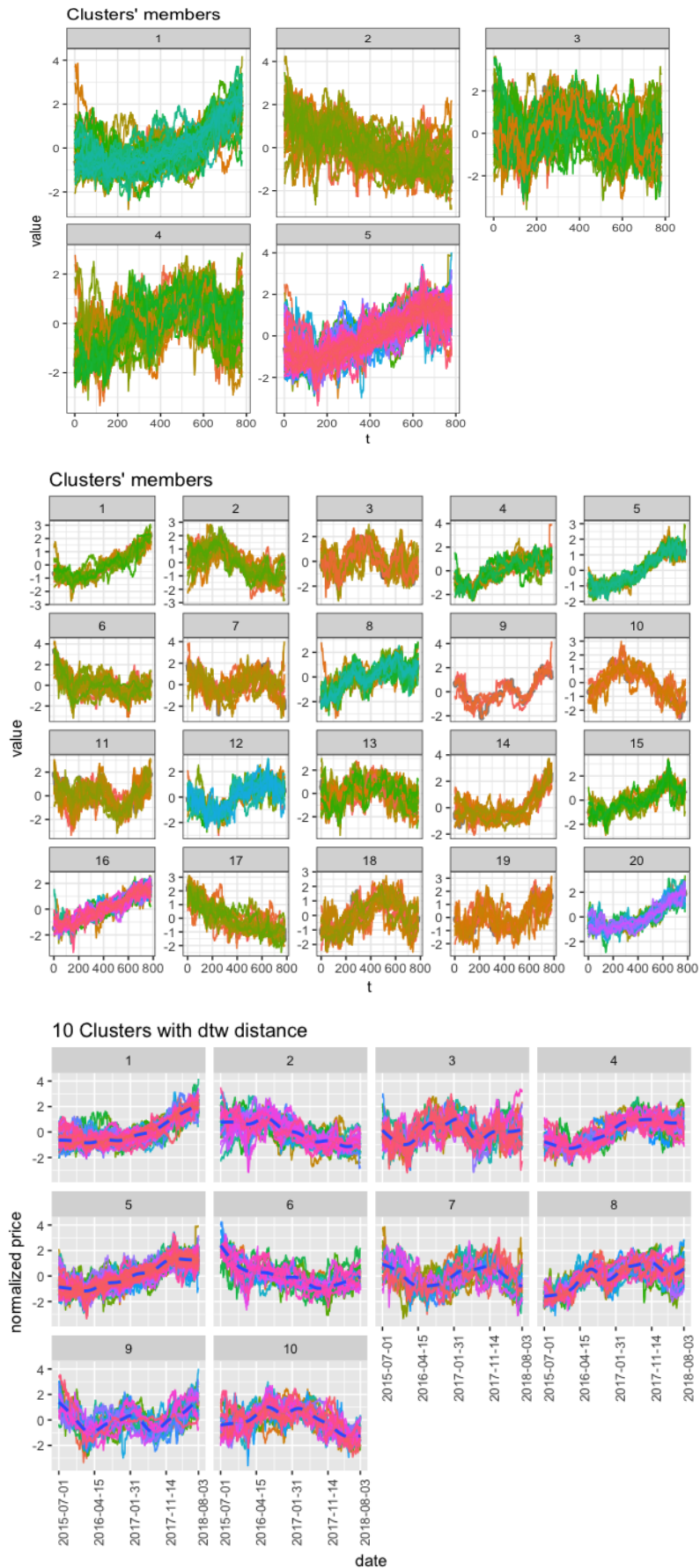
⁴³ (Sardá-Espinosa, 2018) p. 1

⁴⁴ (Ajram, 2020) ps. 68 y 69

⁴⁵ [Stock Clustering with Time Series Clustering in R](#) Visto el 16/10/2020 en: <https://medium.com/@panda061325/stock-clustering-with-time-series-clustering-in-r-63fe1fabe1b6>

⁴⁶ [dtwclust](#), Visto el 16/11/2020 en: <https://cran.r-project.org/web/packages/dtwclust/index.html>

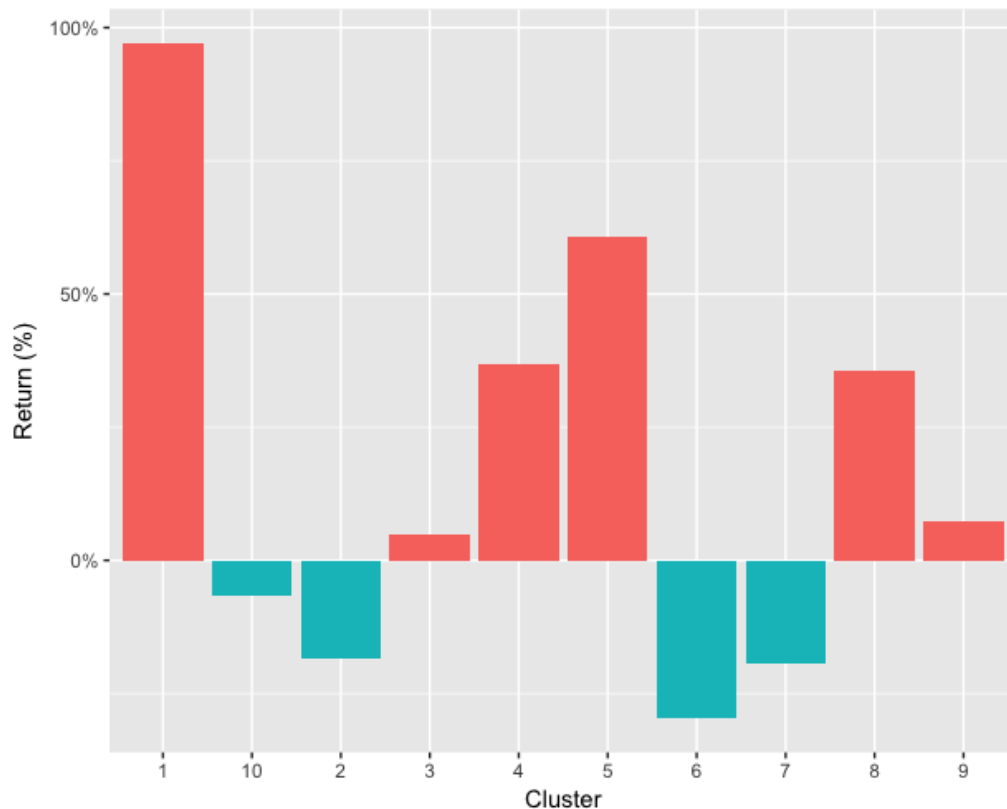
Gráfico 7 Clusterings para $k = 5, 20$ y 10 , respectivamente (Fuente: "[Stock Clustering with Time Series Clustering in R](#)")



Como vemos en el gráfico 7, realiza distintas particiones en $k = 5$, $k = 10$ y $k = 20$ clústeres. Y es que como vimos a la hora de hacer la partición no existe ningún método objetivo que asegure la mejor partición posible para determinar cuál es el número de clústeres óptimo.

Centrando el análisis en $k = 10$ produce una serie de *insights* acerca de cada clúster.

Gráfico 8 Rentabilidad por clústeres (Fuente: "idem")



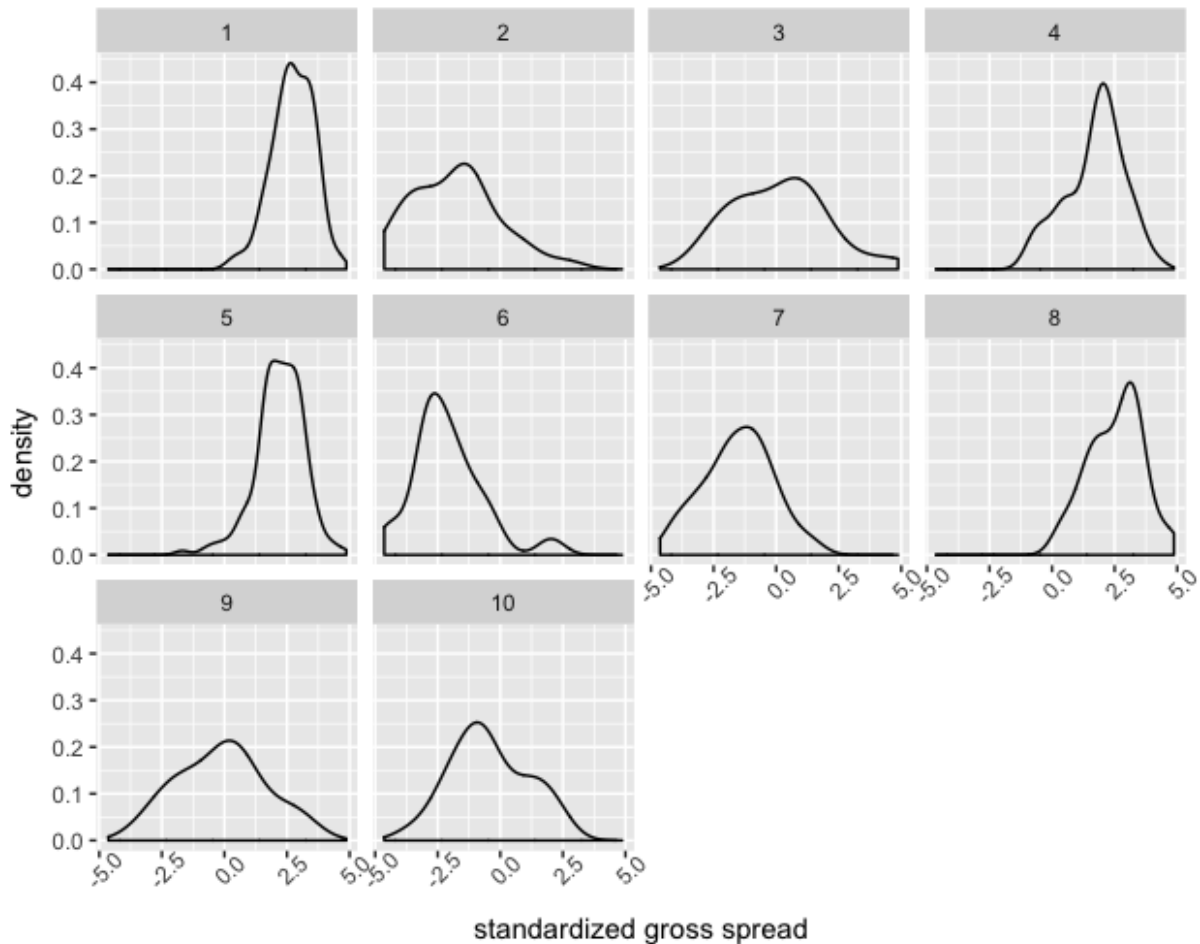
- Como vemos el clúster 1 fue el más lucrativo al ser el que más ha crecido. Si un trader estuviera interesado en invertir su dinero en acciones, y tuviese una cantidad limitada de capital, debería haber invertido en los valores del clúster 1, que son los que más han crecido.
- El clúster 1 asimismo presenta un retorno cercano al 97%, lo que supera ampliamente la rentabilidad del periodo del resto de clústeres. Sin embargo, el peor clúster ofrece un retorno de la inversión del -30%.
- Además, observamos que las posiciones largas en valores del índice son mucho más lucrativas que posiciones cortas, al ser los incrementos mayores que las depreciaciones. La mayoría de los clústeres⁴⁷, 6 frente a 4, se revalorizaron. Respectivamente los clústeres: 1, 3, 4, 5, 8 y 9 (rojos), frente a los: 10, 2, 6 y 7 (azules).

⁴⁷ Otro enfoque interesante sería examinar el número de valores que pertenecen a cada clúster

- Los clústeres positivos no sólo son más, también crecen más que lo que se deprecian los clústeres negativos, sus incrementos son mayores.

Otro análisis intercluster interesante mencionado en el artículo es la volatilidad de cada conjunto, el riesgo.

Gráfico 9 Riesgo asociado a cada clúster (Fuente: "idem")



La interpretación de las Gráfica 9 de arriba nos lleva a considerar que los conjuntos 1, 4, 5, y 8 son aquellos que presentan un menor riesgo al concentrarse sus funciones de densidad de sus retornos en valores positivos.

Como conclusión al análisis clúster, si la tendencia del mercado analizada es extrapolable, el conjunto de valores claramente preferido para cualquier inversor ha de ser el 1, que presenta el mayor ROI⁴⁸ con el menor riesgo asociado.

El análisis clúster tiene una función eminentemente exploratoria, su utilidad reside en jugar con los clústeres a fin de obtener *insights* interesantes para el modelizador.

⁴⁸ Return On Investment (ROI), retorno sobre la inversión

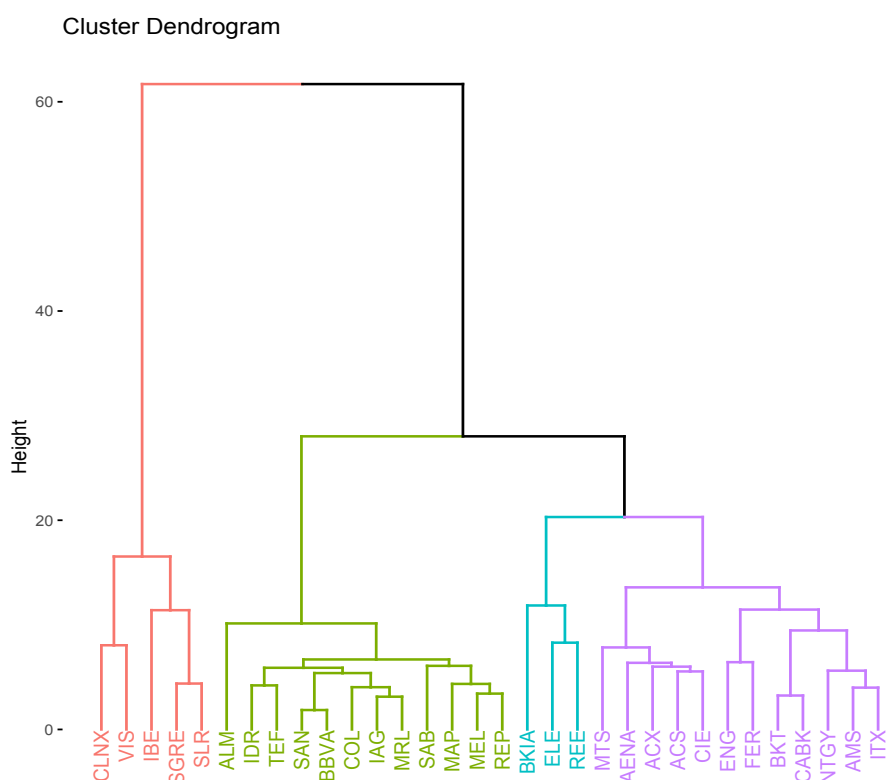
d. Caso práctico

Poniéndonos en la piel de un trader que está interesado en analizar valores del IBEX35 en los que invertir, vamos a poner en práctica los conocimientos expuestos hasta ahora, combinando las diferentes técnicas que hemos aprendido. Para ello utilizando el paquete de R `quantmod`⁴⁹ elaboraremos un *webscraping* de la información OHLC de los distintos valores del IBEX para el periodo comprendido entre el 01/01/2020 hasta el 10/11/2020⁵⁰.

I. Reducción de dimensionalidad: Clustering Jerárquico

Esta técnica del clustering puede ofrecer una primera información interesante al agrupar por los valores por pares de proximidad, creando sucesivamente pares de valores o de conjuntos que se mueven igual. Veamos la siguiente Ilustración 4:

Ilustración 3 Dendrograma del IBEX35 (Fuente: elaboración propia en R)



⁴⁹ [quantmod](https://www.quantmod.com), Visto el 16/11/2020 en: <https://www.quantmod.com>

⁵⁰ Recordando el inicio de este trabajo, las conclusiones de este estudio solo tienen sentido si la información histórica utilizada es extrapolable al futuro, ello requiere entre otras cosas que la información provenga de una misma distribución estadística. Dada la actual crisis del COVID difícilmente se cumple esta condición necesaria, pero, por otro lado, nos permite conocer cómo los distintos valores se han comportado durante la pandemia.

En la Ilustración 4 observamos valores del IBEX35. Como vemos los valores que más se parecen entre sí son Santander (SAN) y BBVA, seguidos de Bankinter (BKT) y Caixabank (CABK), probablemente, bajo las mismas condiciones, factores que afecten a la cotización de uno afectarán igualmente al otro. Luego vemos que Iberdrola (IBE) y Bankia (BKIA) son los valores cuyas fluctuaciones son más distintas a los del resto del índice.

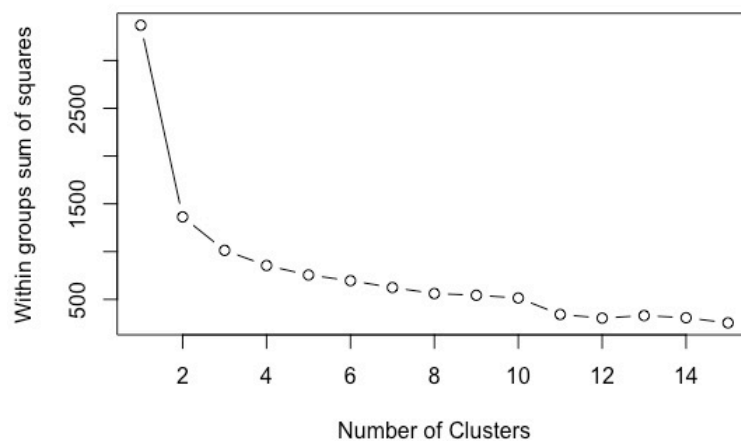
Por colores distinguimos los que más se parecen entre sí, y a su vez qué parecido tienen estos conjuntos con el resto. Así los conjuntos morado y turquesa son los más parecidos mientras que el rojo es con mucha diferencia el menos parecido, con las consiguientes consecuencias que ello implica.

II. Clusterización parcial

Establecida la relación de proximidad en la fluctuación de los distintos valores vamos a agrupar ahora y analizar el comportamiento de distintos grupos de datos al igual que en el análisis del S&P500.

Pero, recordemos que el clusterizar significa crear nuevas etiquetas en las que dividir nuestros datos, correspondiendo cada etiqueta al clúster al que pertenece. Para ello existen diversas métricas que nos ayudan a optimizar el número k , de particiones de nuestro dataset en diferentes clústeres.

Gráfico 10 Suma de los cuadrados para k particiones (Fuente: elaboración propia en R)



En la Gráfica 10 observamos la suma de los cuadrados resultante de dividir nuestro dataset en k particiones. Con este método buscamos minimizar la distancia entre las observaciones pertenecientes a cada grupo, medida en términos cuadráticos. Como resultado deberíamos escoger un número k que minimice el agregado total de cada una de las distancias por grupos. Bajo este método el k óptimo es de 14 y 12 pero recordemos que por simplicidad puede resultar

conveniente utilizar el menor número de particiones posibles, por lo que consideraremos con este método $k = 12$.

Por el contrario, a la hora de clusterizar no solo es importante escoger las particiones k que optimice la similitud intracluster, también hemos de tener en cuenta la conveniencia para su manejo. Puede ser mejor utilizar menos clústeres para sintetizar la información. Por ejemplo, otros dos métodos importantes como CVI (*Cluster Validity Indices*) son el *silhouette* y el *gap statistic*⁵¹. *Silhouette* por su parte determina la calidad de la clusterización, analizando cómo de cercana es la etiquetación a un conjunto de una observación, en otras palabras, su correcta asignación a cada clúster. Por su parte el *gap statistic* maximiza la distancia intercluster, optimiza que los clústeres sean lo más diferenciados posibles.

Gráfico 11 Gap statistic

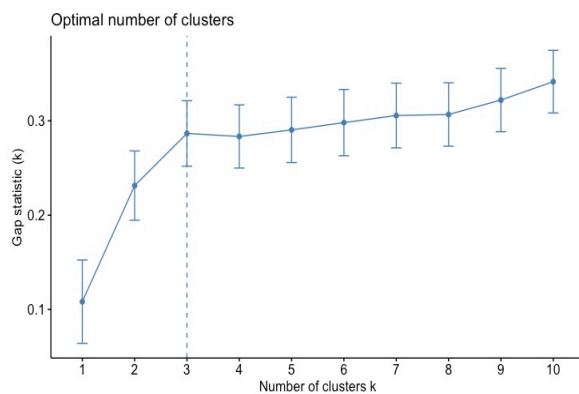
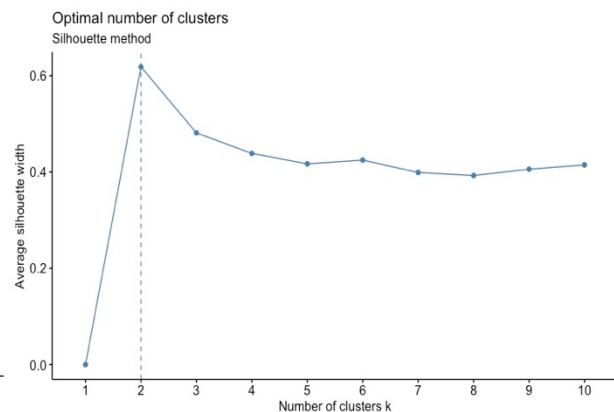


Grafico 12 Silhouette



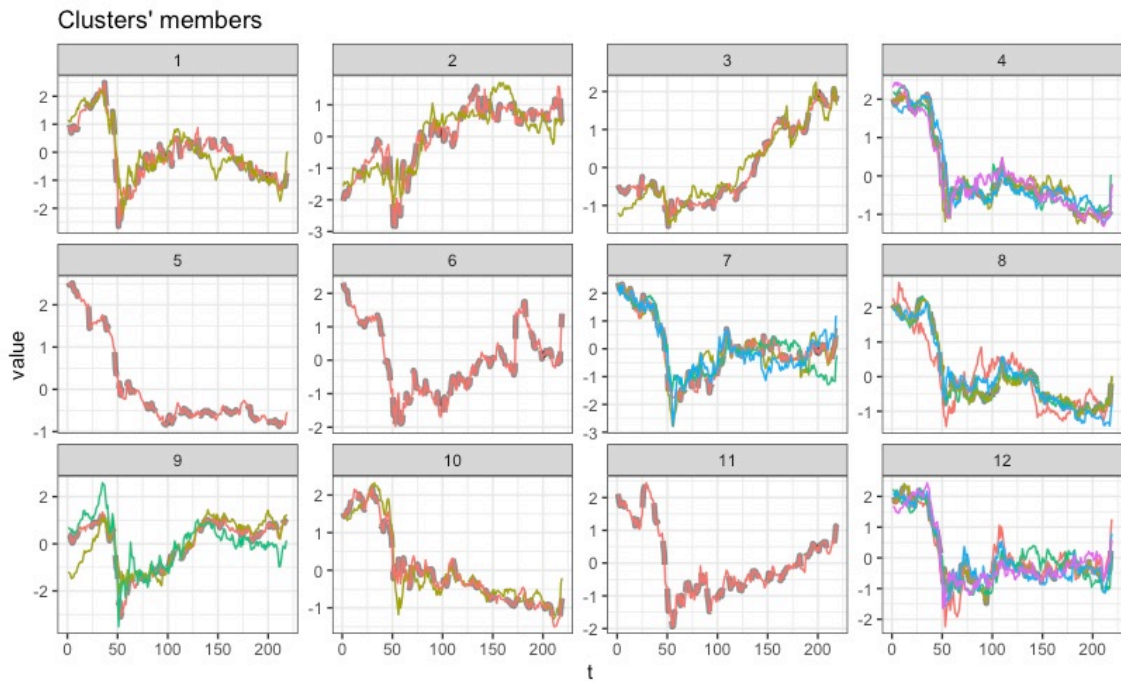
(Fuente: elaboración propia en R)

A continuación, elaboramos la clusterización final con $k = 3, 6$ y 12 (3 y 12 al ser las particiones óptimas según los diferentes métodos y 6 al ser una solución intermedia). Aunque estos métodos nos aseguren que nuestro número de clústeres k sea el recomendable para determinadas condiciones, también nos interesa maximizar la operatividad de la partición.

Los resultados del clustering particional son los siguientes:

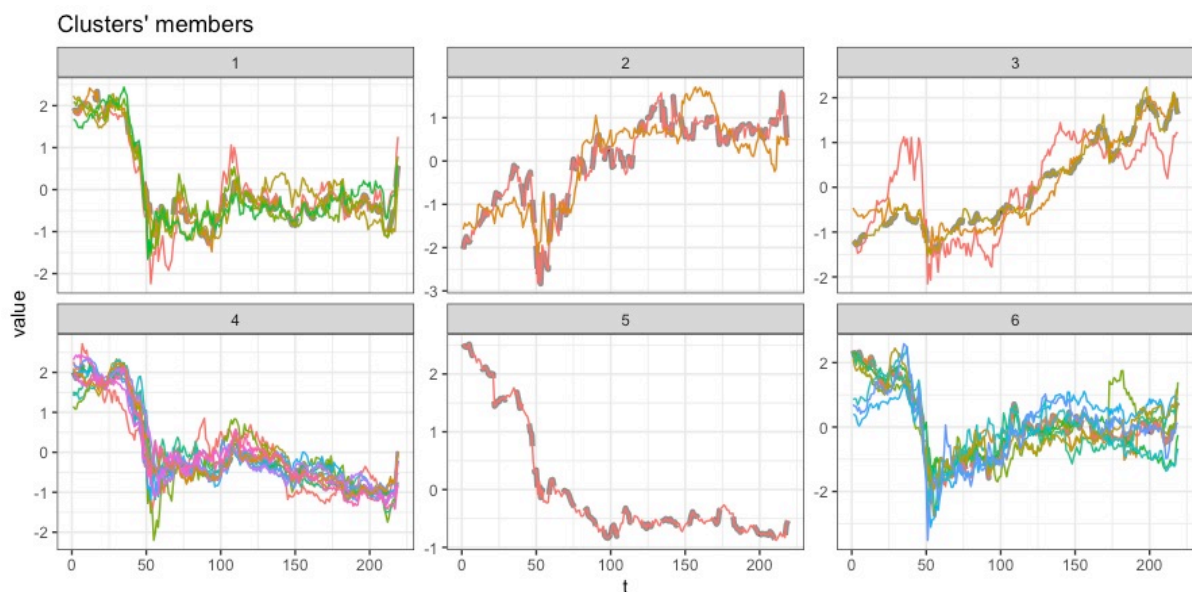
⁵¹ [The complete guide to clustering analysis](https://towardsdatascience.com/the-complete-guide-to-clustering-analysis-10fe13712787) Visto el 16/11/2020 en: <https://towardsdatascience.com/the-complete-guide-to-clustering-analysis-10fe13712787>

Gráfico 12 Clustering para $k = 12$ (Fuente: elaboración propia en R)



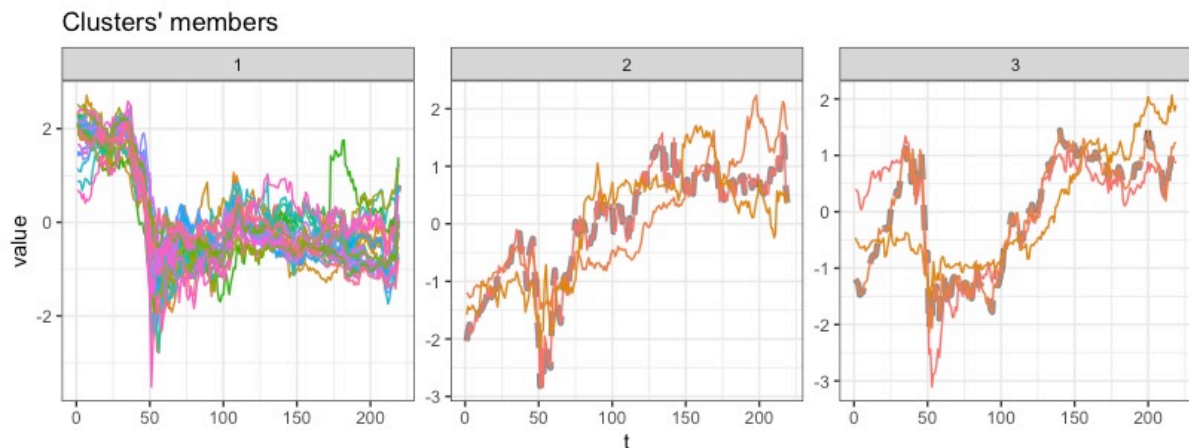
Para $k = 12$ observamos que la fluctuación de algunos clústeres es muy parecida entre sí (12, 11, 10, 8 y 4), destacando como los valores que mejor han reaccionado durante el 2020 a la crisis de la pandemia son los pertenecientes al clúster 3 (liderado por Cellnex y Viscofan), siendo los peores los del clúster 5 (liderado por Sabadell). Como vemos, todos han sufrido caídas importantes a partir del día 50 de cotización, esto es en marzo de 2020. De cualquier modo, esta clusterización sigue siendo demasiado grande para ser manejable, clústeres como el 4 y el 8 apenas presentan diferencias sustanciales.

Gráfico 13 Clustering para $k = 6$ (Fuente: elaboración propia en R)



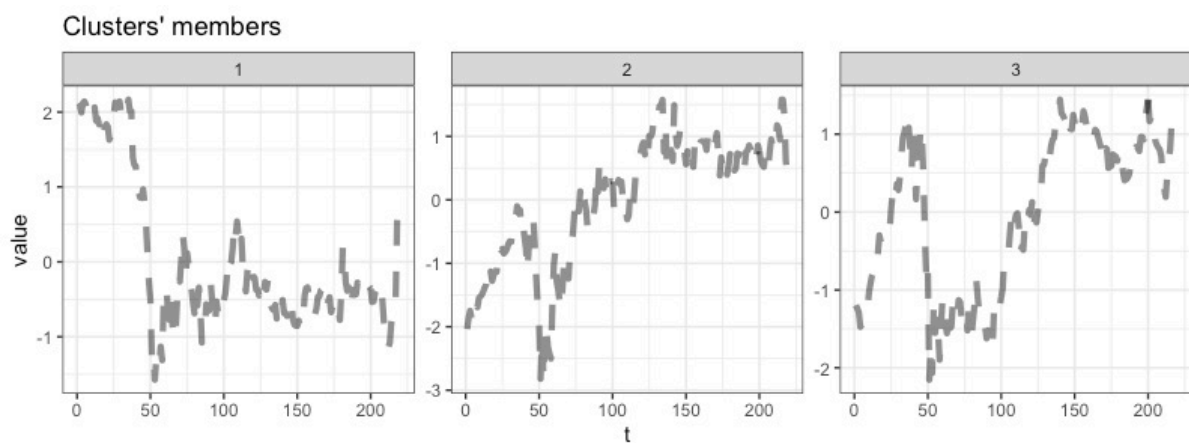
La partición en $k = 6$ nos es útil para simplificar los clústeres, pero sin embargo observamos que clústeres como el 3 hacen una mala agrupación según la monotonicidad, la fluctuación del valor rojo es mayor y distinta a la del valor amarillo, por ejemplo.

Gráfico 14 Clustering para $k = 3$ (Fuente: elaboración propia en R)



Para limpiar la imagen que nos ofrece la partición $k = 3$ podemos simplificar los clústeres en sus prototipos usando sus centroides. Pero un primer análisis nos indica que debido a la gran gama de color y la hiperdensificación el clúster 1 es el que más valores recoge, siendo el 2 y el 3 las excepciones.

Gráfico 15 Prototipos de cada clúster (Fuente: elaboración propia en R)



Ahora, distinguimos nítidamente cual es el comportamiento entre clústeres. El 1 ha sufrido una enorme depreciación, sin haberse aún recuperado, y el 2 y el 3 se han recuperado, siendo más pronta la recuperación en el 2 que entre $t = (50, 100)$ ya se había producido un efecto rebote en sus cotizaciones.

Como conclusión de este análisis clúster observamos que los activos más interesantes a ojos de un trader, dada la coyuntura de esta etapa, son precisamente aquellos menos expuestos al riesgo de nuevas oleadas de contagios, y las respectivas consecuencias sociales y económicas. Una opción barajaría abrir posiciones cortas en los valores cuya fluctuación más vaya a depender en el empeoramiento de la pandemia, pero al abrir posiciones cortas estaríamos asumiendo ganancias potenciales limitadas junto pérdidas potenciales ilimitadas. Todo ello en un entorno de alta incertidumbre, y por tanto de alta volatilidad en las cotizaciones.

Desde una perspectiva de inversor prudente resulta más atractivo invertir en aquellos valores “inmunes” a la pandemia abriendo posiciones largas, que por contra limitan las pérdidas, pero no limitan las ganancias.

En consecuencia, nos interesan valores como Viscofan (VIS), Cellnex (CLNX) y Solaria (SLR), a la sazón, la empresa del IBEX35 que más ha se ha revalorizado desde el inicio del 2020⁵².

III. Regresión

Vistos lo valores que menos exposición tienen a la pandemia vamos a intentar pronosticar mediante el aprendizaje no supervisado utilizando un modelo ARIMA⁵³ (*Auto Regressive Integrated Moving Average*) en R, utilizado comúnmente para manejar series temporales estacionarias, esto es que no presentan una media, covarianza y autocovarianza estática, característica esta que facilitan su tratamiento estadístico.

Es bastante frecuente en datos de demanda o precios de acciones que sean no estacionarios, pero distinguiendo dentro de las series podemos conseguir que lo sean, o que se acerquen a serlo.

⁵² [IBEX35](https://www.expansion.com/mercados/cotizaciones/indices/ibex35_I.IB.html), Visto el 17/11/2020 en: https://www.expansion.com/mercados/cotizaciones/indices/ibex35_I.IB.html

⁵³ [ARIMA](https://rpubs.com/valeamasso/386527), Visto el 18/11/2020 en: <https://rpubs.com/valeamasso/386527>

Gráfico 16 Regresión con ARIMA de Solaria (Fuente: elaboración propia en R)

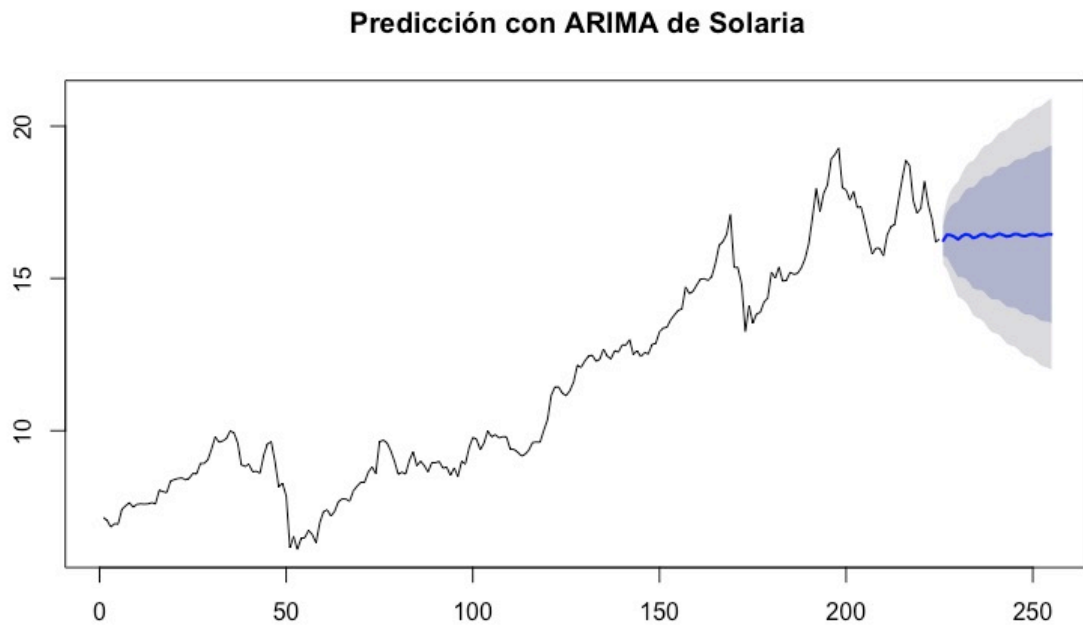


Gráfico 17 Regresión con ARIMA de Cellnex (Fuente: elaboración propia en R)

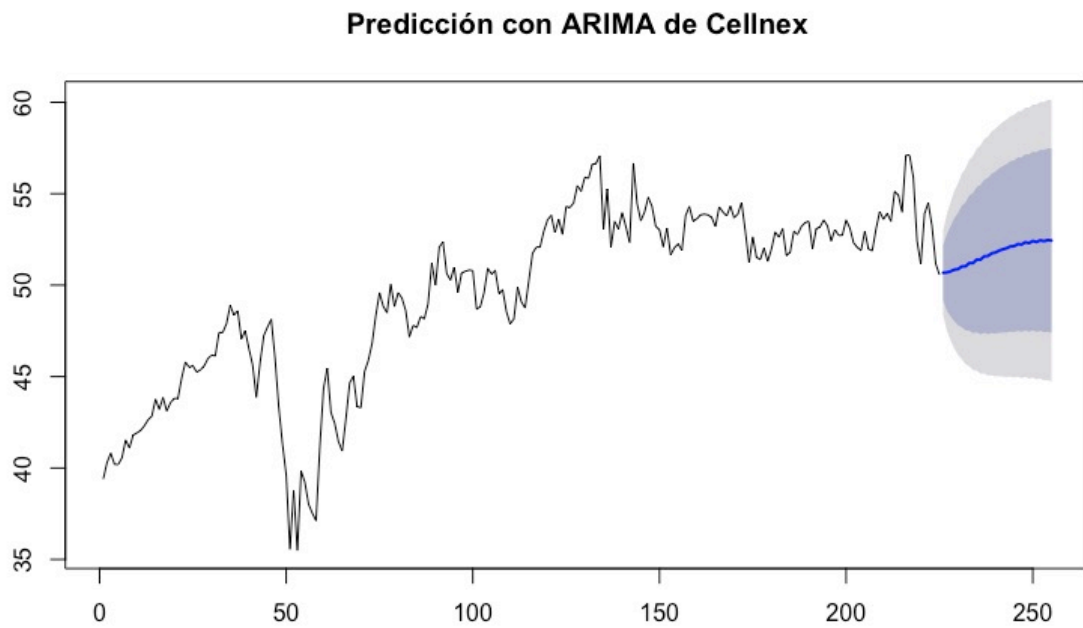
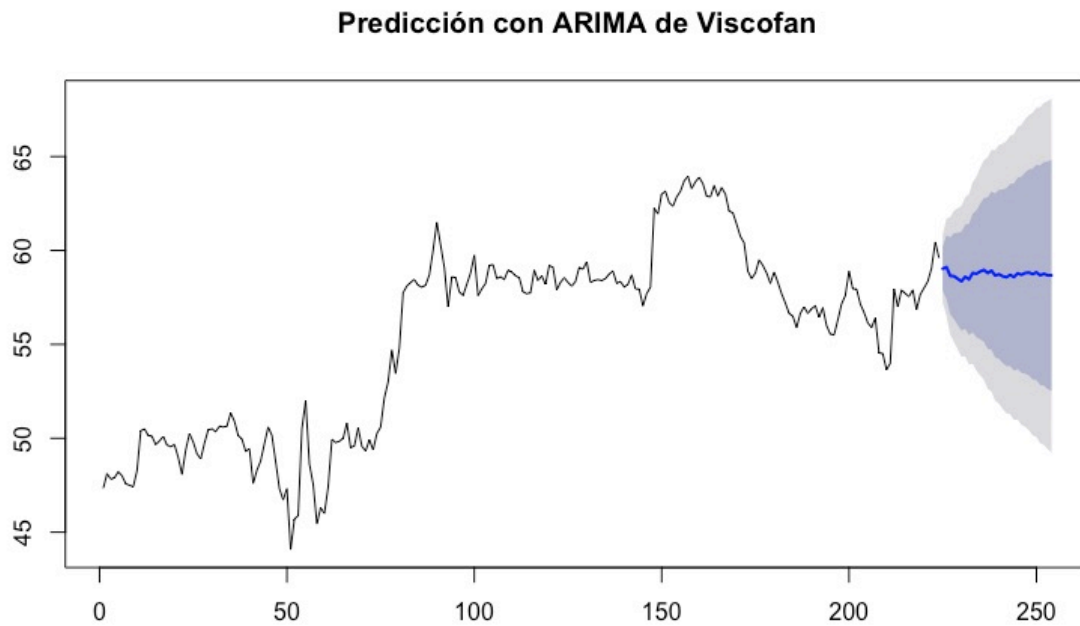


Gráfico 18 Regresión con ARIMA de Viscofan (Fuente: elaboración propia en R)



Las gráficas 16, 17 y 18 anteriores nos ofrecen una regresión a 30 días del valor en la línea azul mostrada, con un intervalo de confianza del 80% entorno a las áreas sombreadas oscuras y otro del 95% entorno a las áreas grises, más amplias. Como se observa, según más nos alejamos del momento actual mayor es la volatilidad de la predicción, los intervalos se expanden a lo largo del eje vertical.

Utilizando modelos ARIMA, vemos que el activo del que mejor rendimiento se espera es Cellnex, que muestra una predicción claramente ascendente. ¿Qué nos muestra el modelo del comportamiento esperado de este valor?

La conclusión a la que se llega es la siguiente:

Tabla 4 Intervalos de confianza de Cellnex (Fuente: elaboración propia en R y Excel)

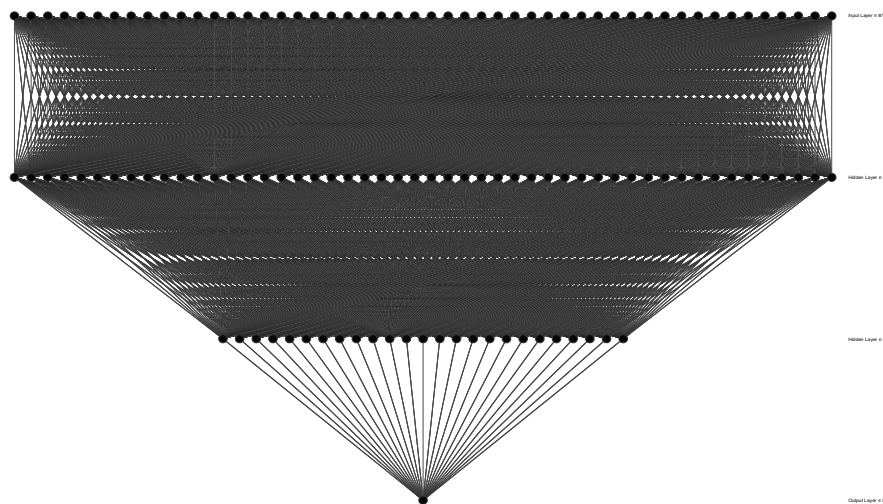
día futuro	valor predicho	Intervalo de Confianza					
		80%			95%		
		mín.	máx.	amplitud	mín.	máx.	amplitud
1	50,5326	48,9541	52,1110	3,1569	48,1186	52,9466	4,8281
2	50,7040	48,5990	52,8090	4,2100	47,4847	53,9233	6,4386
...
29	52,1914	45,2272	59,1557	13,9286	41,5405	62,8424	21,3019
30	52,2410	45,1502	59,3318	14,1815	41,3966	63,0854	21,6887

Centrándonos en la Tabla 3, en el intervalo del 95%, vemos que en el peor de los casos podemos esperar que la cotización de Cellnex en 30 días llegue a los 41 euros, y en el mejor de los casos a los 63 euros.

Podemos realizar adicionalmente un modelo más complejo de Deep learning en Python, utilizando redes neuronales con el modelo LSTM (Long Short Term Memory), al igual que ARIMA utilizado para series temporales, para contrastar que resultados obtenemos con este otro método, desde otro punto de vista.

El diseño de la red neuronal consta de 4 capas, las dos primeras primeras de 50 neuronas, la tercera de 25 y la ultima siendo la que nos devuelve el output deseado, la regresión.

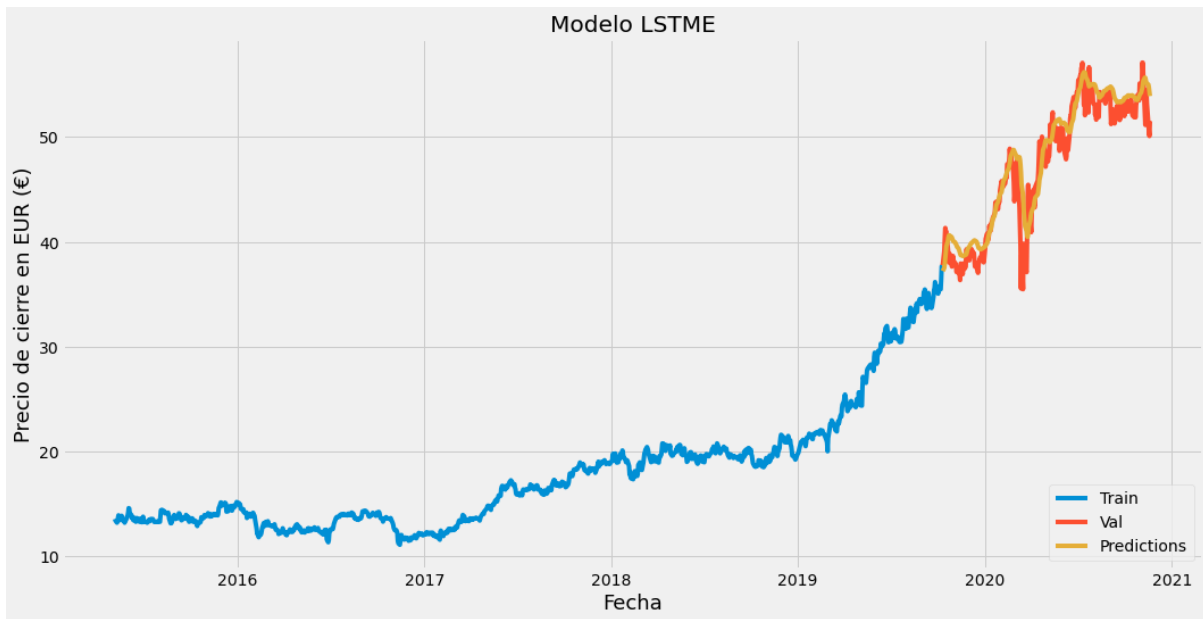
Ilustración 4 Esquema de la Red Neuronal (Fuente: elaboración propia)



La regresión devuelta por nuestro modelo es la siguiente, utilizando como conjunto de entrenamiento los precios de cierre del periodo 5-YTD (*Years To Date*), vamos a utilizar como inputs los precios, normalizados (no estandarizados) en $[0,1]$ para mejor asimilación por parte los perceptrones, de los últimos 60 días para predecir el precio futuro.

Normalizar es de los pocos ajustes que requieren de la ingeniería de características para modelos de Deep learning.

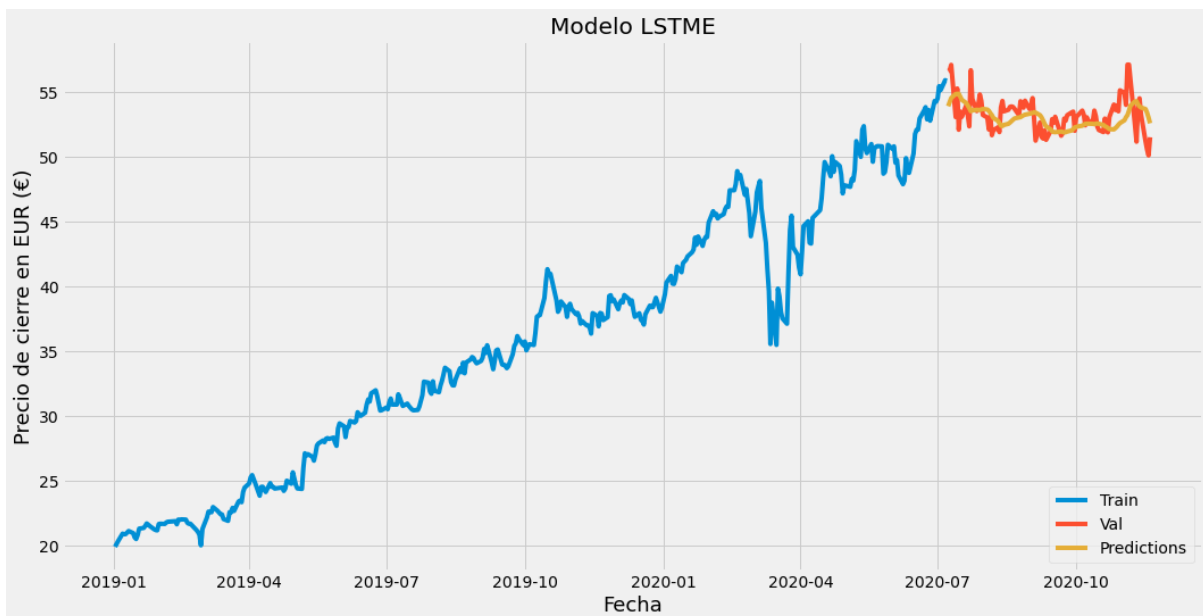
Gráfico 19 Red Neuronal para el periodo 2015-2020 (Fuente: elaboración propia en Python)



La Gráfica 19 sin duda parece mostrar una predicción bastante fidedigna del comportamiento real de Cellnex, parece incluso predecir la caída de la acción durante los primeros meses del CoVID, empezando en marzo de 2020, aunque la caída real fuese mayor del doble que la predicha.

Pero observemos de cerca los resultados del mismo modelo para un periodo 2-YTD:

Gráfico 20 Red Neuronal para el periodo 2019-2020 (Fuente: elaboración propia en Python)



Vemos en el Gráfico 20, que el error de nuestro modelo a corto plazo es mayor. Las predicciones se solapan e interseccionan menos con menos datos, y se distancian más de los valores reales. Yerran más. Esto no es de extrañar dadas dos circunstancias: la alta volatilidad en periodos de alta incertidumbre como el estudiado, y la tendencia al *overfitting* del machine learning, que aplicado fuera del conjunto de validación tiende a fallar.

Como mayor aportación de la red neuronal LSTM vemos que es capaz de pronosticar mejor los cambios de tendencia. Ello lleva a pensar que para un futuro experimento exploratorio la red neuronal sería más útil como clasificador que como regresor.

e. Interpretación general de los resultados obtenidos

Para finalizar esta parte 4, cabe recordar que la estadística y la modelización, como se advertía al principio tan solo son métodos para tratar de comprender realidades de una naturaleza sumamente compleja, y que como consecuencia de ello siempre van a estar sujetas a un error. Además de que corremos el riesgo de que un *outlier* tire por la borda toda nuestra investigación. Aunque fuésemos capaces de comprender plenamente el comportamiento de un fenómeno en base a su historia, no quita que en el futuro surjan otros factores desconocidos que somos incapaces de imaginar, que alteren la distribución de nuestros datos.

*Caveat emptor*⁵⁴.

El análisis técnico, el chartismo, la estadística, y el uso del machine learning, resumidos en este trabajo suponen avances históricos en la inversión, que aportan valor añadido. Pese a todo, sería un disparate construir cualquier estrategia de inversión olvidando otros aspectos básicos, especialmente el análisis fundamental, y el riesgo sistémico y de mercado, que por las limitaciones requeridas no han sido profundizados.

Aun así, el machine learning y la IA tienen una gran variedad de usos potenciales en el análisis de valores, pudiendo crear modelos de predicción de demanda, o de evolución de costes, o de riesgo esperado.

En definitiva, las técnicas vistas en esta parte 4, no son la herramienta ni única ni exclusiva ni la más fiable, pero sí las más avanzadas. Sobre todo en un campo de nuestra realidad tan dominado por la incertidumbre como son los mercados, como conjunto de interacciones humanas en las que intervienen innumerables factores.

⁵⁴ El “comprador”, utilizador, de estas técnicas ha de ser consciente de que existe un riesgo inherente a su compra.

Podemos extraer unas conclusiones válidas de nuestros análisis, pero unos estándares mínimos para una buena praxis deben incluir: contrastarlas; si se puede, refutarlas; y en especial complementarlas; contra otros análisis y modelos, y sobretodo contra la lógica.

5. Conclusiones

Como hemos visto la revolución tecnológica, va a empapar todas las facetas de nuestra vida, hacia una progresiva algoritmización, en la que el ser humano va a alcanzar un nivel cognoscitivo superior, al ser capaz de asimilar cantidades inmensamente mayores de información.

No solo eso, la cuantificación de la información supone desprendernos de los sesgos subjetivos de nuestra consciencia. Ello significa que podemos elaborar algoritmos independientes a las tendencias subjetivas a las que normalmente nos inclinamos.

Los algoritmos, como hemos visto, son capaces de aproximar fenómenos tan complejos como los mercados, en los que los innumerables agentes y factores que influyen en ellos hacen imposible la extracción de un axioma matemático unificador, con el que entenderlos como un conjunto. La pura economía en sus conceptos más simples tradicionalmente estudiados, como la ley de la oferta y la demanda, se construye sobre los tambaleantes modelos de aproximaciones en los puntos de equilibrio, que pese a ser muy cercanas a la realidad, son sin embargo relaciones imperfectas.

Construyendo modelos de machine learning tampoco somos capaces de alcanzar esa aspiración de entender la ley universal por la que se rige el comportamiento humano, ni las interacciones sociales, ni por tanto la economía y los mercados; pero sí de hacer una aproximación sustancialmente mejor a los métodos preexistentes. Mejor en términos de rendimiento, velocidad y accesibilidad; resumidos en procesos automatizados.

En los próximos años veremos cómo el machine learning, como la punta de lanza de la IA, la algoritmización y robotización de nuestro entorno van a suponer el mayor avance de la humanidad en años. Si en la actualidad los modelos de machine learning utilizados primordialmente son modelos del tipo supervisados y no supervisados, de *shallow learning*, estos tan solo suponen los modelos más simples que podemos construir.

Pero, como vimos, con un simple modelo de clasificación podemos hacer un análisis exploratorio para conocer el movimiento del precio de una criptomoneda, con un acierto de el 98%. Sin duda supone un ejemplo práctico de la aplicabilidad y utilidad de los modelos de machine learning al trading financiero.

Con todo, el Deep learning y el aprendizaje reforzado, recordemos que son respectivamente el aprendizaje sobre varias capas y el aprendizaje que tiene en cuenta el entorno y el comportamiento de otros agentes, son actualmente muy utilizados, pero están aun en fase de desarrollo. Queda mucho rango de mejora, especialmente para estos tipos de machine learning.

Tienen un potencial inmensamente mayor y en base a ellos se avanzará exponencialmente hasta alcanzar probablemente la singularidad, entendida por Raymond Kurzweil, inventor y director de ingeniería de Google, como el momento en el que la IA sobrepase a la inteligencia humana. Y no deberíamos sorprendernos, como hemos visto el fondo cuántico de Renaissance Technologies es capaz de batir, por el doble, la rentabilidad del mejor inversor humano. En el otro ejemplo más cercano sería desde luego impensable, a priori, que cualquier humano fuese capaz de lograr un acierto del 98% en los movimientos diarios de cualquier valor.

Lo cierto es que la algoritmización y el uso de modelos de machine learning son desde hace tiempo el pan de cada día de la inversión y los mercados financieros, de ahí que, como hemos visto, los organismos reguladores se apuren para regular su uso como vimos en el trading algorítmico.

Este proceso va a empapar con los años el futuro de nuestras vidas.

Pese a las amenazas que presenta su implementación en nuestra sociedad: destrucción de empleos, desigualdad en el reparto de beneficios, exclusión social, posiciones privilegiadas, etc.; inexorablemente el camino de la humanidad se dirige a alcanzar la singularidad. Y, si realmente queremos alcanzar un conocimiento más allá de lo que es humanamente posible hemos de prepararnos, adaptarnos y trabajar en pos su llegada, ya que también acarrea consecuencias positivas como el alargamiento de la vida y la mejora de su calidad.

Pero para que toda la humanidad se beneficie necesitamos democratizar la tecnología y el conocimiento, algo que durante los últimos años estamos afortunadamente viviendo gracias al libre mercado y el desarrollo de los recursos online.

El Big Data, el análisis de datos y el machine learning solo son un pequeño píxel en el gran cuadro del porvenir de la humanidad. Pero son, sin embargo, al igual que: el desarrollo de la capacidad de computación, las técnicas de computación cuántica, los avances matemáticos y otros descubrimientos que nos deparan; un pequeño pero necesario píxel para completar la imagen de nuestro futuro.

6. Bibliografía

a. Artículos académicos

- Madan, I., Saluja, S., Zhao, A. (2014): [Automated bitcoin trading via machine learning algorithms](#), Stanford University, Stanford. *GitHub*. En: <https://github.com/safwankdb/Machine-Learning-Project-Ideas/blob/master/Automated%20Bitcoin%20Trading%20via%20Machine%20Learning%20Algorithms.Isaac%20Madan%2C%20Shaurya%20Saluja%2C%20Aojia%20Zhao.pdf>
- Sardá-Espinosa, A. (2019): [Comparing Time-Series Clustering Algorithms in R Using the dtwclust Package](#). *The R Journal* 11:1, pages 22-43. En: <https://journal.r-project.org/archive/2019/RJ-2019-023/index.html>

b. Libros

- Ajram, J. (2020): *Estrategias de operativa bursátil*, Deusto, Barcelona.
- Burkov, A. (2019): *The hundred-page machine learning book*. <http://themlbook.com/>
- Chollet, F. (2019): *Deep Learning con Python*, Anaya, Madrid.
- Costa Ran, L., Font Vilalta, M. (1993): *Commodities: mercados financieros sobre materias primas*, Esic, Madrid.
- Lacalle, D. (2013): *Nosotros los mercados*, Deusto, Barcelona.
- Lopez Zafra, J. M., Queralt, R. (2019): *Alquimia, cómo los datos se están transformando en oro*, Deusto, Madrid.
- Rankia (2016): *Manual de análisis fundamental y valoración de empresas*, Rankia.

c. Recursos Web

- [AI-Powered Fund Thrashing S&P 500 Is Wary of Cyclical Bounce](#). Visto el 25/10/2020 en: <https://www.bloomberg.com/news/articles/2020-10-21/market-beating-robot-fund-manager-shys-away-from-cyclical-bets>
- [Algorithmic Trading](#). Visto el 20/09/2020 en: <https://www.investopedia.com/terms/a/algorithmictrading.asp#:~:text=Algorithmic%20trading%20is%20a%20process,to%20the%20market%20over%20time.>
- [ARIMA](#), Visto el 18/11/2020 en: <https://rpubs.com/valeamasso/386527>
- [BBVA ¿Qué es el 'trading'?](#) Visto el 14/06/2020 en: <https://www.bbva.com/es/que-es-trading-que-hace-falta-para-operar/>

- [El fondo que convierte 100 dólares en 400 millones en 30 años.](https://cincodias.elpais.com/cincodias/2019/11/15/mercados/1573831472_956898.html) Visto el 21/09/2020 en:
https://cincodias.elpais.com/cincodias/2019/11/15/mercados/1573831472_956898.html
- [Flash Trading](https://www.investopedia.com/terms/f/flash-trading.asp), Visto el 16/10/2020 en: https://www.investopedia.com/terms/f/flash-trading.asp
- [IBEX35](https://www.expansion.com/mercados/cotizaciones/indices/ibex35_I.IB.html), Visto el 17/11/2020 en: https://www.expansion.com/mercados/cotizaciones/indices/ibex35_I.IB.html
- [Inside the JPMorgan Trading Desk the U.S. Called a Crime Ring](https://www.bloomberg.com/news/articles/2020-09-28/inside-the-jpm-precious-metals-desk-called-a-crime-ring-by-prosecutors), Visto el 28/09/2020 en: https://www.bloomberg.com/news/articles/2020-09-28/inside-the-jpm-precious-metals-desk-called-a-crime-ring-by-prosecutors
- [quantmod](https://www.quantmod.com), Visto el 16/11/2020 en: https://www.quantmod.com
- [Stock Clustering with Time Series Clustering in R](https://medium.com/@panda061325/stock-clustering-with-time-series-clustering-in-r-63fe1fabe1b6) Visto el 16/10/2020 en: https://medium.com/@panda061325/stock-clustering-with-time-series-clustering-in-r-63fe1fabe1b6
- [The complete guide to clustering analysis](https://towardsdatascience.com/the-complete-guide-to-clustering-analysis-10fe13712787) Visto el 16/11/2020 en: https://towardsdatascience.com/the-complete-guide-to-clustering-analysis-10fe13712787

d. Legislación e informes de autoridades oficiales

- Staff of the Division of Trading and Markets U.S. Securities and Exchange Commission (2014): *Equity Market Structure Literature Review Part II: High Frequency Trading*.
- DIRECTIVA 2014/65/UE DEL PARLAMENTO EUROPEO Y DEL CONSEJO de 15 de mayo de 2014 relativa a los mercados de instrumentos financieros y por la que se modifican la Directiva 2002/92/CE y la Directiva 2011/61/UE.

7. Anexos

a. R code

I. Webscraping de las empresas del IBEX35

```
# WEB SCRAPING
library('rvest')

## Loading required package: xml2

url = 'https://es.wikipedia.org/wiki/IBEX_35'

webpage = read_html(url)
webpage

## {html_document}
## <html class="client-nojs" lang="es" dir="ltr">
## [1] <head>\n<meta http-equiv="Content-Type" content="text/html; charset
=UTF-8 ...
## [2] <body class="mediawiki ltr sitedir-ltr mw-hide-empty-elt ns-0 ns-su
bject ...

IBEX35 <- url %>%
  html() %>%
  html_nodes(xpath = '//*[@id="mw-content-text"]/div[1]/table[3]') %>%
  html_table()

## Warning: 'html' is deprecated.
## Use 'xml2::read_html' instead.
## See help("Deprecated")

IBEX35 <- IBEX35[[1]]

head(IBEX35)

##   Ticker      Empresa      Sede Entrada
## 1   ANA          Acciona Alcobendas  2015
## 2   ACX          Acerinox   Madrid    2015
## 3   ACS          Grupo ACS   Madrid    1998
## 4   AENA         Aena       Madrid    2015
## 5   ALM          Almirall   Barcelona 2020
## 6   AMS Amadeus IT Group Madrid    2011
##                                     Sector[13]      ISIN Ponderación
## 1                                     Construcción ES0125220311      1,20
## 2      Mineral, metales y transformación ES0132105018      0,44
## 3                                     Construcción ES0167050915      1,45
## 4                                     Transporte y distribución ES0105046009      3,96
## 5 Productos farmacéuticos y biotecnología ES0157097017      0,38
## 6                                     Electrónica y software ES0109067019      4,52

IBEX35_tickers= IBEX35$Ticker
```


II. Modelo ARIMA

```
# ARIMA Models
library(tseries)
library(timeSeries)
library(forecast)
library(xts)
library(quantmod)

getSymbols(c('VIS.MC', 'CLNX.MC', 'SLR.MC'), from='2020-01-01', adjusted=TRUE)
class(VIS.MC) # analizamos individualmente los tres valores de arriba
datos<-VIS.MC[,4]%>%na.omit
plot(datos)

par(mfrow=c(1,2))
Acf(datos, main='ACF para series diferenciadas')
Pacf(datos, main='PACF para series diferenciadas')

print(adf.test(datos)) # hipotesis nula es no estacionaridad
auto.arima(datos, seasonal = FALSE)

fitA = auto.arima(datos, seasonal = FALSE)
tsdisplay(residuals(fitA), lag.max = 80, main = '(0,1,0) Model residuals')
auto.arima(datos, seasonal = FALSE) # necesitamos BIC y AIC que sea lo minimo posible

fitB=arima(datos, order = c(1,2,4)) # custom ARIMA
tsdisplay(residuals(fitB), lag.max = 40, main = '(1,2,4) Model residuals')

fitC=arima(datos, order = c(5,1,4))
tsdisplay(residuals(fitC), lag.max = 40, main = '(5,1,4) Model residuals')

fitD=arima(datos, order = c(1,1,1))
tsdisplay(residuals(fitD), lag.max = 40, main = '(1,1,1) Model residuals')

# Plots of ARIMA models

par(mfrow=c(2,2))
term<-30 # periodos, depende del numero de dias que empleemos

fcast1<-forecast(fitA, h=term)
plot(fcast1)

fcast2<-forecast(fitB, h=term)
plot(fcast2)

fcast3<-forecast(fitC, h=term)
plot(fcast3)

fcast4<-forecast(fitD, h=term)
plot(fcast4)
```

```
# MAPE (Mean Accuracy of the Percent Error), 100 - MAPE = Accuracy
```

```
accuracy(fcast1)  
accuracy(fcast2)  
accuracy(fcast3)  
accuracy(fcast4)
```

```
fcast2
```

III. Clustering con DTW y Dendograma

```
# DTW
```

```
IBEX35_empresas = c("ANA.MC" , "ACX.MC" , "ACS.MC" , "AENA.MC" , "ALM.MC" ,  
"AMS.MC" , "MTS.MC" ,  
"SAB.MC" , "SAN.MC" , "BKIA.MC" , "BKT.MC" , "BBVA.MC" , "  
CABK.MC" , "CLNX.MC" ,  
"CIE.MC" , "ENG.MC" , "ELE.MC" , "FER.MC" , "GRF.MC" , "  
IAG.MC" , "IBE.MC" ,  
"ITX.MC" , "IDR.MC" , "COL.MC" , "MAP.MC" , "MEL.MC" , "M  
RL.MC" , "NTGY.MC" ,  
"PHM.MC" , "REE.MC" , "REP.MC" , "SGRE.MC" , "SLR.MC" , "  
TEF.MC" , "VIS.MC" )  
getSymbols( IBEX35_empresas , from= "2020-01-01", adjusted=TRUE)
```

```
IBEX <- cbind (ANA.MC$Adjusted , ACX.MC$ACX.MC.Adjusted , ACS.MC$ACS.MC.Ad  
justed , AENA.MC$AENA.MC.Adjusted ,  
ALM.MC$ALM.MC.Adjusted , AMS.MC$AMS.MC.Adjusted, MTS.MC$MTS  
.MC.Adjusted ,  
SAB.MC$SAB.MC.Adjusted , SAN.MC$SAN.MC.Adjusted , BKIA.MC$BKIA.M  
C.Adjusted, BKT.MC$BKT.MC.Adjusted ,  
BBVA.MC$BBVA.MC.Adjusted , CABK.MC$CABK.MC.Adjusted, CLNX.MC$CLNX  
.MC.Adjusted,  
CIE.MC$CIE.MC.Adjusted , ENG.MC$ENG.MC.Adjusted , ELE.MC$ELE.MC.  
Adjusted , FER.MC$FER.MC.Adjusted ,  
GRF.MC$Adjusted , IAG.MC$IAG.MC.Adjusted , IBE.MC$IBE.MC.Adjuste  
d ,  
ITX.MC$ITX.MC.Adjusted , IDR.MC$IDR.MC.Adjusted , COL.MC$COL.MC.  
Adjusted, MAP.MC$MAP.MC.Adjusted , MEL.MC$MEL.MC.Adjusted ,  
MRL.MC$MRL.MC.Adjusted , NTGY.MC$NTGY.MC.Adjusted ,  
PHM.MC$Adjusted , REE.MC$REE.MC.Adjusted , REP.MC$REP.MC.Adjuste  
d ,SGRE.MC$SGRE.MC.Adjusted ,  
SLR.MC$SLR.MC.Adjusted , TEF.MC$TEF.MC.Adjusted ,VIS.MC$VIS.MC.A  
djusted )
```

```
head(IBEX)  
anyNA(IBEX)  
SIN_NAs<-na.omit(IBEX)  
anyNA(SIN_NAs)
```

```
SIN_NAs_normalizado<-scale(SIN_NAs)  
SIN_NAs_normalizado_T<-t(SIN_NAs_normalizado)
```

```

rownames(SIN_NAs_normalizado_T)<-c( "ACX" , "ACS" , "AENA" , "ALM" , "AMS"
, "MTS" , "SAB" , "SAN" , "BKIA" , "BKT" , "BBVA" ,
      "CABK" , "CLNX" , "CIE" , "ENG" , "ELE" ,
      "FER" , "IAG" , "IBE" , "ITX" , "IDR" , "COL" , "M
AP" , "MEL" , "MRL" , "NTGY" ,
      "REE" , "REP" , "SGRE" , "SLR" , "TEF"
      , "VIS" )
df<-SIN_NAs_normalizado_T

dtw_cluster6 = tsclust(df, type="partitional",k=6,
                      distance="dtw_basic",centroid = "pam",seed=1234,trace=T,
                      args = tsclust_args(dist = list(window.size = 5)))

plot(dtw_cluster6,type='centroids')

dtw_cluster3 = tsclust(df, type="partitional",k=3,
                      distance="dtw_basic",centroid = "pam",seed=1234,trace=T,
                      args = tsclust_args(dist = list(window.size = 5)))

plot(dtw_cluster3,type='centroids')

dtw_cluster3@clusinfo

dtw_cluster12 = tsclust(df, type="partitional",k=12,
                      distance="dtw_basic",centroid = "pam",seed=1234,trace=T,
                      args = tsclust_args(dist = list(window.size = 5)))

plot(dtw_cluster12, type="centroids",clus=1L)

plot(dtw_cluster12, )

dtw_cluster12@clusinfo

# Silhouette method
fviz_nbclust(df, kmeans, method = "silhouette") +
  labs(subtitle = "Silhouette method")

# CLUSTERING

# Datos:
df<-SIN_NAs_normalizado_T

# Modo 1:
# Distancias
res.hc <- hclust(dist(df))
dev.new(); fviz_dist(dist(df))

# Dendrograma

```

```

dev.new(); fviz_dend(res.hc)
#dev.new(); fviz_dend(res.hc, rect = TRUE, cex = 0.5)

# Modo 2:
# Distancias y divide en k clusters
res <- hcut(df, k = 4) #res < hcut(USArrests, k = 4, stand = TRUE)
# Dendrograma
dev.new(); fviz_dend(res)
dev.new(); fviz_dend(res, rect = TRUE, cex = 0.5, k_colors = c("red", "blue", "#E7B800", "green"))

```

IV. Charting del oro

```

# Chartismo
library(quantmod)
getSymbols("XAU/USD",src="oanda") # Gold from Oanda
head(XAUUSD)

chartSeries(XAUUSD)
addMACD() # Moving Average Convergence Divergence
addBBands() # Bollinger bands
addRSI() # Relative Strnght Index
addSMI() # Stochastical momentum index

```

b. Python code

- *# Artificial Recurrent Neural Network called LSTM (Long Short Term Memory), # predicting closing stock price using past 60 days price*

```

import math
import pandas_datareader as web
import numpy as np
import pandas as pd
from sklearn.preprocessing import MinMaxScaler
from keras.models import Sequential
from keras.layers import Dense, LSTM
import matplotlib.pyplot as plt
plt.style.use('fivethirtyeight')

#Get the stock quote
df = web.DataReader('CLNX.MC', data_source='yahoo', start='2015-01-01')
#Show the data
df

# Tensor info
df.shape

# Visualize data

plt.figure(figsize=(16,8))
plt.title('Precio de cierre historico')
plt.plot(df['Close'])
plt.xlabel('Date', fontsize=18)

```

```

plt.ylabel('Precio de cierre en EUR (€)', fontsize=18)
plt.show()

#Create a new dataframe with only the 'Close' column
data = df.filter(['Close'])
#Converting the dataframe to a numpy array
dataset = data.values
#Get /Compute the number of rows to train the model on
training_data_len = math.ceil( len(dataset) *.8)

training_data_len # 80% del conjunto de entrenamiento

#Scale the all of the data to be values between 0 and 1
scaler = MinMaxScaler(feature_range=(0, 1))
scaled_data = scaler.fit_transform(dataset)

scaled_data

#Create the scaled training data set
train_data = scaled_data[0:training_data_len , : ]
#Split the data into x_train and y_train data sets
x_train=[]
y_train = []
for i in range(60, len(train_data)):
    x_train.append(train_data[i-60:i,0])
    y_train.append(train_data[i,0])
    if i<=60:
        print(x_train)
        print(y_train)
        print()

#Convert x_train and y_train to numpy arrays
x_train, y_train = np.array(x_train), np.array(y_train)
x_train.shape

#Reshape the data into the shape accepted by the LSTM
x_train = np.reshape(x_train, (x_train.shape[0],x_train.shape[1],1))
x_train.shape
# np is 2D but LSTM expects it to be 3D (time steps, number of observa
tions and number of features)

#Build the LSTM network model
model = Sequential()
model.add(LSTM(units=50, return_sequences=True, input_shape=(x_train.sh
ape[1],1))) # 1st Layer, x_train.shape = time steps (=60), 50 neurons
model.add(LSTM(units=50, return_sequences=False)) # 2nd Layer, 50 neur
ons
model.add(Dense(units=25)) # 3rd Layer, 25 neurons
model.add(Dense(units=1)) # 4th Layer

#Compile the model
model.compile(optimizer='adam', loss='mean_squared_error') # we set as
a loss function MSE

```

```

#Train the model
model.fit(x_train, y_train, batch_size=1, epochs=1) # fit= to train, B
atch (Lote) size = length of the Batch
#epoch = when number of iterations when an entire dataset is passed th
rough a Neural Network

#Test data set
test_data = scaled_data[training_data_len - 60: , : ]
#Create the x_test and y_test data sets
x_test = []
y_test = dataset[training_data_len : , : ] #Get all of the rows from
index 1603 to the rest and all of the columns
#(in this case it's only column 'Close'), so 2003 - 1603 = 400 rows of
data
# y_test are the future values we want to predict
for i in range(60,len(test_data)):
    x_test.append(test_data[i-60:i,0])

#Convert x_test to a numpy array
x_test = np.array(x_test)

#Reshape the data into the shape accepted by the LSTM
x_test = np.reshape(x_test, (x_test.shape[0],x_test.shape[1],1))
# Number of samples = x_test.shape[0],
# Number of time steps = x_test.shape[1],
# Number of features = 1

#Getting the models predicted price values
predictions = model.predict(x_test)
predictions = scaler.inverse_transform(predictions)#Undo scaling

#Calculate/Get the value of RMSE
rmse=np.sqrt(np.mean(((predictions- y_test)**2)))
rmse

#Plot/Create the data for the graph
train = data[:training_data_len]
valid = data[training_data_len:]
valid['Predictions'] = predictions
#Visualize the data
plt.figure(figsize=(16,8))
plt.title('Modelo LSTME')
plt.xlabel('Fecha', fontsize=18)
plt.ylabel('Precio de cierre en EUR (€)', fontsize=18)
plt.plot(train['Close'])
plt.plot(valid[['Close', 'Predictions']])
plt.legend(['Datos de entrenamiento', 'Valor real', 'Predicciones'], 1
oc='lower right')
plt.show()

```