



Facultad Ciencias Económicas y Empresariales

MACHINE LEARNING APLICADO AL TRADING

Autor: Macarena Salvador Maceira

Director: José Portela González

MADRID | Junio 2019

Índice

I.	Resumen y Abstract.....	6
1.	Introducción.....	8
1.1	Contexto del Big Data:.....	10
1.1.1	Definición de Big Data	11
1.1.2	Dimensiones del Big Data	12
1.1.3	Fuentes del Big Data	15
1.1.4	Cadena de Valor del Big Data	16
1.1.5	Tipos de análisis	18
1.1.6	Aplicaciones en la actualidad.....	19
2.	Machine Learning:	22
2.1	Tipos de aprendizaje.....	23
2.1.1	Aprendizaje supervisado.....	23
2.1.2	Aprendizaje no supervisado	25
2.1.3	Otros tipos de aprendizaje	26
2.2	Deep Learning y redes neuronales	27
2.2.1	Perceptrón multicapa	30
2.2.2	Redes recurrentes	38
3.	Machine Learning en Trading:.....	41
4.	Conclusiones	52
5.	Bibliografía	53

Índice de ilustraciones

Ilustración 1 Nube de palabras con los principales términos cuando se habla de Big Data	11
Ilustración 2 Cadena de valor del Big Data	16
Ilustración 3 Paradigma de la programación tradicional	22
Ilustración 4 Paradigma de Machine Learning	22
Ilustración 5 Proceso de desarrollo del aprendizaje supervisado	24
Ilustración 6 Técnicas de análisis del aprendizaje supervisado	25
Ilustración 7 Técnicas de análisis del aprendizaje no supervisado	26
Ilustración 8 Perceptrón simple	31
Ilustración 9 Gráficas de problemas resueltos mediante funciones a partir de un perceptrón simple o varios perceptrones simples	32
Ilustración 10 Perceptrón simple con función de activación	33
Ilustración 11 Esquema de un perceptrón multicapa (MLP)	35
Ilustración 12 Representación de la función de Back-propagation en un perceptrón multicapa	37
Ilustración 13 Representación de la función dropout en un perceptrón multicapa	37
Ilustración 14 Esquema de redes LSTM con una o varias memory blocks	39
Ilustración 15 Esquema de un memory block y sus componentes principales	40
Ilustración 16 Niveles de riesgo de activos financieros según la legislación española	45

Índice de tablas

Tabla 1 Unidad de medida de los datos	13
Tabla 2 Algoritmos de aprendizaje más importantes	29
Tabla 3 Principales funciones de activación	34
Tabla 4 Tipos de activos disponibles en los Mercados financieros	42
Tabla 5 Ratings de los activos financieros en función de diferentes agencias de rating	44
Tabla 6 Herramientas que proporcionan mejores resultados en predicción de precios en mercados financieros	49

I. Resumen y Abstract

RESUMEN

Con el auge del Big Data en la actualidad están surgiendo nuevas formas y herramientas de análisis, como es el Machine Learning, que ayudan a una mejora en la eficacia y eficiencia de la toma de decisiones.

Este trabajo realiza una síntesis a cerca de múltiples estudios realizados sobre las aplicaciones de Machine Learning en el mundo de las finanzas, centrándose en la actividad de trading. Para ello he realizado una investigación de la literatura académica especializada sobre aplicaciones de *Machine Learning*. De entre todas las herramientas de *Machine Learning* se destaca especialmente el uso de las redes neuronales artificiales (RNA), con el fin de comprender cual es la razón del uso de estas herramientas para la labor de análisis de los mercados financieros y predicción de precios.

Como resultado de analizar y comparar las diferentes RNA utilizadas en la labor de predicción de los precios de los diferentes activos financieros se concluye que las redes que mejores resultados ofrecen son las LSTM, ya que tienen en cuenta la variable tiempo.

Palabras clave: Machine Learning, Proyección de series temporales, Redes neuronales artificiales, Trading, Perceptrón multicapa, redes LSTM.

ABSTRACT

In the last two decades there has been an increase in the use and importance of Big Data and the tools and different forms of analysis that it brings. One of those tools that is gaining popularity is Machine Learning, which leads to a more efficient and effective way of solving problems and decision making.

The aim of this study is to synthesize multiple studies about Machine Learning applications in finance, focusing on trading and the uses it has for its work. For this purpose, I had carried out a review of the specialized literature about the applications of Machine Learning with artificial neural networks (ANN) as the main tool of this forecasts, with the purpose of having a full understanding about the reasons of using these tools.

After comparing the different ANN used for forecasting the prices of financial assets, I've concluded that the best results came from the possibility of taking time into account, given by the LSTM networks.

Key words: Machine Learning, Forecasting time series, Artificial neural networks, Trading, Multilayer perceptron, LSTM networks.

1. Introducción

Este trabajo recoge una extensa investigación sobre las técnicas de *Machine Learning* y algunas de sus herramientas más utilizadas. Estas técnicas han ido ganando importancia en los últimos años gracias a los avances tecnológicos y el uso del Big Data, que permite encontrar soluciones más simples a problemas del día a día.

Especialmente me voy a centrar en intentar explicar de la manera más sencilla posible cómo funcionan las redes neuronales, una herramienta compleja que cada vez cobra más importancia en el entorno de *Machine Learning*, y en sus aplicaciones.

Dentro del sector financiero se pueden utilizar para resolver diversos problemas y como herramienta de apoyo en la toma de decisiones, como puede ser el análisis de riesgos financieros, los patrones de blanqueo de dinero o la evasión de impuestos.

Una de las aplicaciones que más importancia está tomando en los últimos años, y en la que yo he decidido centrar este estudio, es en trading, ya sea de manera automática o como una herramienta de análisis para los inversores. El éxito en este campo es debido no solo a la gran capacidad matemática y analítica que tienen estas herramientas, sino a su capacidad para hacer predicciones. Estas predicciones son la clave para el *high frequency trading*, posible gracias a estas herramientas y en las actividades especulativas, que reducen su riesgo sin tener que renunciar al posible beneficio.

Gracias al Big Data, estas herramientas han ido tomando cada vez más importancia dentro del mundo financiero, aunque no son tan recientes. Existen múltiples estudios acerca de cuales son las mejores herramientas para la predicción de las tendencias y los precios en los mercados financieros.

En ellos proponen el uso de herramientas más simples como puede ser un árbol de decisión o SVM (*Support Vector Machine*), que se utilizarían como herramientas de clasificación para elegir los activos financieros que puedan dar mayor beneficio deducido a partir de sus características. O herramientas complejas como el uso de *Deep Learning*

y redes neuronales artificiales que son capaces de realizar predicciones muy ajustadas a la realidad.

El fin de este trabajo es conseguir comprender el funcionamiento de algunas de las herramientas más complejas que se utilizan en trading, sintetizando y resumiendo diferentes estudios que se han hecho en los últimos años. En ellos proponen el uso de aprendizaje reforzado, Ritter (2017), en el que se proporciona un modelo de recompensas cuya finalidad es conseguir el mayor beneficio posible realizado mediante un *Modelo de decisión de Markov*, el uso de redes neuronales artificiales con pequeñas diferencias en su arquitectura y funciones de activación, propuesto por Ramos y Martínez (2013), Moghaddam, Moghaddam y Esfandyari, (2016), Atsalakis y Valavanis (2009), el uso de *Deep Learning*, propuesto por Heaton, Polson y Witte (2016), y Navon y Keller (2017) o el uso de Redes neuronales de *Deep Learning* para el *high frequency trading*, propuesto por Ganesh y Rakheja (2018).

Para ello comenzaré por introducir qué es el *Big Data* y cuales son sus características más destacables, debido a que el éxito de las herramientas de *Machine Learning* viene dado por la facilidad del acceso a grandes cantidades de datos. A continuación, explicaré qué es el Machine Learning y cuáles son las características generales de los tipos de aprendizaje que existen, y me centraré en las herramientas más aplicadas en el análisis de mercados y trading. Por último, explicaré las aplicaciones en trading y comentaré cuales son las herramientas con las que se obtienen mejores resultados.

La metodología llevada a cabo para conseguir estos objetivos ha sido una extensa labor de investigación. Durante este proceso de investigación he leído diferentes publicaciones académicas en las que se explicaba el funcionamiento de las herramientas de *Machine Learning* más utilizadas en la predicción de los precios de los activos financieros y las correcciones que eran necesarias para obtener la predicción mas real posible.

Para conseguir un mayor entendimiento sobre la razón para utilizar esas herramientas determinadas y no otras, he investigado las diferentes herramientas de *Machine Learning*

utilizadas para hacer predicciones, y las correcciones necesarias para evitar problemas que surgen debido al uso de herramientas muy complejas.

Debido a la gran variedad de herramientas de Machine Learning utilizadas en este campo, y las limitaciones que conlleva este trabajo, he decidido centrarme en la aplicación de redes neuronales artificiales y Deep Learning. La razón por la que he escogido estas herramientas ha sido para poder realizar un estudio acerca de si, teniendo en cuenta su grado de dificultad, son las herramientas que otorgan mejores resultados.

1.1 Contexto del Big Data:

Actualmente vivimos en una sociedad digital. La mayoría de las actividades diarias de las personas están conectadas a la red generando datos, ya sea de manera directa como puede ser el uso de las redes sociales dando a “like” a publicaciones, o de manera indirecta, usando las tarjetas de transporte público en Madrid cada vez que viajas en metro. Esta generación masiva de datos es la que muchos consideran el centro de la “Cuarta Revolución Industrial”, que generará un cambio en la manera de enfocar los negocios y la creación de riqueza, destacando la digitalización y automatización de procesos.

Esta tendencia creciente de generación de datos se lleva observando durante las dos últimas décadas, y las previsiones de crecimiento no disminuyen, sino que aumentan con el paso del tiempo y el desarrollo de la tecnología.

Ya lo decía John Naisbitt: “*Nos estamos ahogando en información y estamos hambrientos de conocimiento*”¹ en su libro *Megatrends* (1982). El problema al que nos enfrentamos mayoritariamente con la gran cantidad de información generada es el uso que darle y como almacenarla de manera correcta. Por lo que el uso de *Big Data* y sus herramientas ayudan a reducir estos problemas y generar valor a través de los datos.

¹ Cita original en inglés: “We are drowning in information and starving for knowledge.” *Megatrends: Ten New Directions Transforming Our Lives* (1982).

1.1.1 Definición de Big Data

Cuando se habla de Big Data no hay una definición exacta a la que todos los académicos y los que utilizan los datos hagan referencia de la misma manera. El término dependiendo del contexto, se utiliza para hacer referencia, por ejemplo, a la cantidad de datos generados o a los procesos de transformación que han de sufrir esos datos para sacar información útil para las empresas.

La definición más aceptada para el término Big Data es el *Modelo de las 3Vs*, publicado por Doug Laney (2001), aunque en un principio no se propuso como una definición del término, ahora es ampliamente utilizado y son las características básicas del *Big Data*.

Por ello, De Mauro, Greco y Grimaldi (2014), realizaron un estudio que buscaba crear una definición consensuada sobre el *Big Data*, analizando con herramientas de *Text mining* las palabras más utilizadas en artículos y estudios sobre el *Big Data* y el análisis de datos. Para ello estudiaron las palabras más repetidas en más de 1500 artículos y destacaron que palabras se repetían más y cual era la información más relevante en ellos.

Ilustración 1 Nube de palabras con los principales términos cuando se habla de Big Data



Fuente: De Mauro, Greco y Grimaldi, (2014)

Como podemos comprobar por la nube de palabras generada a partir del análisis de los diferentes artículos, las palabras más destacadas después de “*Big data*”, son información, las 3Vs (volumen, velocidad y variedad), análisis, almacenamiento y técnicas entre otras. En su estudio también estudian otros aspectos comunes de los artículos como la información a la que se refieren, la tecnología utilizada, los métodos de transformación y el impacto del uso de *Big Data* en las empresas y nuestras vidas.

Por ello su definición final sobre *Big Data* tiene en cuenta las 3Vs que describen las características de la información a la que hace referencia, la tecnología y los métodos usados para transformar los datos y la transformación que crea el valor de estos datos. De manera que definen *Big Data* como:

“El Big Data representa los activos de información caracterizados por un alto volumen, velocidad y variedad que requieren de una tecnología específica y de métodos analíticos para transformarla en valor.”²

Esta definición hace referencia no solo a los datos que forman el *Big Data*, si no a todo el proceso que necesita realizarse para que conforme valor mas allá que el proporcionado por todo el conjunto de datos en crudo.

1.1.2 Dimensiones del Big Data

Como ya he mencionado anteriormente, las características del *Big Data* se determinan por el *Modelo de las 3Vs* propuesto por Laney al que se le han ido añadiendo más dimensiones a medida que se ha ido extendiendo el *Big Data*.

² Cita original en inglés: “*Big Data represents the Information assets characterized by such a High Volume, Velocity and Variety to require specific Technology and Analytical Methods for its transformation into Value.*”

- Volumen: hace referencia a la cantidad de datos generados que necesitan ser recolectados y almacenados para su posterior análisis. Con el paso del tiempo y el incremento en el uso de dispositivos que generan datos se han pasado de Gigabytes a Terabytes, y la unidad de medida sigue en aumento. En el año 2009 Facebook Inc. comunicaba que llegaba a 1 Petabyte y que esperaban un crecimiento exponencial y Google contaba en 2016 con 15 Exabytes de datos.

Tabla 1 Unidad de medida de los datos

Nombre	Símbolo	Potencias binarias y valores decimales
byte	b	$2^0 = 1$
Kbyte	KB	$2^{10} = 1\ 024$
Megabyte	MB	$2^{20} = 1\ 048\ 576$
Gigabyte	GB	$2^{30} = 1\ 073\ 741\ 824$
Terabyte	TB	$2^{40} = 1\ 099\ 511\ 627\ 776$
Petabyte	PB	$2^{50} = 1\ 125\ 899\ 906\ 842\ 624$
Exabyte	EB	$2^{60} = 1\ 152\ 921\ 504\ 606\ 846\ 976$
Zettabyte	ZB	$2^{70} = 1\ 180\ 591\ 620\ 717\ 411\ 303\ 424$
Yottabyte	YB	$2^{80} = 1\ 208\ 925\ 819\ 614\ 629\ 174\ 706\ 176$

Fuente: Elaboración propia

- Velocidad: hace referencia tanto a la transmisión de datos como a la velocidad de crecimiento. Los *dataset*³ son exponenciales, de manera que cada vez son de mayor tamaño y se generan con más rapidez y mayor diversidad. El tiempo juega un papel especial, ya que las tendencias son muy fugaces y los datos que se generan en un mes pueden no servir para el mes siguiente, por lo que es importante no solo el tiempo de generación sino lo que tardas en trasladar los datos para su análisis.

³ Dataset: palabra con la que se define a un conjunto de datos.

- Variedad: se refiere a los diferentes tipos de datos que conforman el *Big Data*. Debido a que los datos pueden surgir de diferentes fuentes, no son todos del mismo tipo (videos, imágenes, palabras o números, por ejemplo) y la manera de almacenarlos y analizarlos varía. Por ello se clasifican en tres grupos: estructurados, semi-estructurados y desestructurados. Los más fáciles de trabajar son los estructurados, aunque conforman el grupo mas pequeño de datos, el 90% de los datos generados son de tipo desestructurado.

Dependiendo de la fuente que uses para definir las dimensiones del *Big Data* se añaden más Vs al modelo, aunque las tres mencionadas anteriormente siempre están presentes.

Destacan:

- Veracidad: hace referencia a la fiabilidad de los datos, si los datos son erróneos o no son precisos, las conclusiones del análisis serán inexactas, por lo que es importante comprobar la veracidad, aunque al tratarse de grandes cantidades de datos es una tarea muy complicada.
- Valor: los datos en sí no aportan valor, pero una vez se analizan con las herramientas adecuadas se consiguen *insights* que sí aportan valor. Además, también hace referencia a la minimización de costes necesarios para adquirir y almacenar los datos, así como su análisis.
- Variabilidad: hace referencia a los cambios que pueden sufrir los datos ya sea con el paso del tiempo o dependiendo del contexto en que se usen, una misma palabra puede tener diferente significado dependiendo de la situación en la que se presente. Es posible que un dato fuese cierto en un cierto momento, por ejemplo, la cantidad de clientes que tiene una empresa en un determinado mes, pero que cambie con el paso del tiempo.
- Visualización: es importante encontrar formas de transmitir la información obtenida a través del análisis de los datos de manera que sea sencillo de comunicar los resultados. Por eso, la visualización es una de las dimensiones más

importantes, ya que, si no se comunica bien, las decisiones tomadas pueden ser erróneas.

Muchos académicos también consideran que hay que añadirle una dimensión extra que sería la C de Complejidad, ya que a la hora de tratar los datos hay que entender bien cual es la fuente, que significan en su contexto y como he de analizarlos para conseguir el mayor beneficio posible.

1.1.3 Fuentes del Big Data

Hay muchas maneras diferentes de clasificar las fuentes de datos, podría ser por el sector que las ha generado (finanzas, ciencia, redes sociales, sanidad, etc.), por el lugar geográfico donde se originan, o por los dispositivos que los originan. Ninguna de las maneras de clasificar las fuentes es perfecta como clasificación, por lo que a continuación voy a explicar cuales son las fuentes que más datos generan.

- Internet de las cosas (IoT⁴): uno de los mayores contribuidores a la generación de datos del *Big Data*. Es el término con el que se denominan a los aparatos conectados a la red, como son los dispositivos móviles, los *wearables*⁵ o los aparatos inteligentes para hogares como *Alexa* de Amazon o *HomePod* de Apple.
- Datos científicos: recogidos a lo largo del tiempo. Hacen referencia a datos sísmicos, del océano, del clima o de astronomía. Actualmente la recogida de estos datos se puede hacer en menores intervalos de tiempo, casi momentáneos.
- Computación en la nube: diferentes servicios y plataformas disponibles online, a disposición del cliente y sus necesidades. Se recogen, almacenan y analizan los datos por lo que generan mas datos.
- Webs y Redes sociales: generados generalmente por los usuarios. Cada vez existen mas sitios web disponibles para los usuarios, que al utilizarlos ya sea para

⁴ IoT: Denominación de internet de las cosas de su nombre en inglés Internet of things.

⁵ Wearables: dispositivos que se llevan puestos y monitorizan datos (Ej. Relojes inteligentes)

realizar alguna transacción como para buscar información generan datos. El uso de Redes Sociales, cada vez más presente en la vida de las personas, genera millones de datos por segundo, que analizados de manera correcta permiten la obtención de mucha información no disponible a simple vista.

En general, independientemente de la fuente, los datos por si solos no generan valor, es necesario analizarlos y combinarlos de manera correcta para así poder realizar una buena estrategia empresarial.

1.1.4 Cadena de Valor del Big Data

La Cadena de Valor es un concepto implantado por Porter (1980), y hace referencia a una serie de procesos de transformación que son necesarios que siga una empresa para generar valor para sus clientes. Esto se puede aplicar al marco en el que se desarrolla el *Big Data* para crear valor a partir de los datos en crudo que es capaz de recoger.

Ilustración 2 Cadena de valor del Big Data



Fuente: Adaptad de Bhadani y Jothimani, (2014)

- 1) Generación de datos: el primer paso cuando se habla de *Big Data* sería la creación de los datos que van a formar los datasets a partir de los cuales se va a realizar el análisis. Como he explicado en el apartado anterior hay muchas fuentes que generan datos.

- 2) Recolección de los datos: en este punto se recogen los datos de todas las fuentes posibles. Los métodos más utilizados son los sensores, rastreadores web, herramientas de monitorización web y se pueden conseguir datos manualmente con data mining.
- 3) Transmisión de los datos: una vez que los datos han sido recolectados, hay que trasladarlos al lugar donde van a ser almacenados. Hay dos tipos de flujo de información *Inter-Dynamic Circuit Network* e *Intra-Dynamic Circuit Network*.
- 4) Preprocesamiento de los datos: los datos recolectados en crudo pueden estar repetidos, ser inconsistentes y contener mucho ruido, por lo que antes de almacenarlos se realizan una serie de análisis con el fin de almacenar solo la información realmente válida. Esto se hace no solo para conseguir información que contenga mayor calidad, sino para ahorrar en costes de almacenamiento de datos.
- 5) Almacenamiento de los datos: debe de ser una plataforma fiable y de acceso fácil, que además sea capaz de soportar las cantidades de datos con las que se trabaja. Los almacenes de datos deberían seguir la *Teoría CAP* propuesta por Brewer (2000). Esta teoría propone que un buen almacenamiento de datos tiene como características la consistencia, la disponibilidad y la tolerancia al particionado (si algún dato falla que no falle todo).
- 6) Análisis de los datos: dependiendo del tipo de datos del que se trate, y de cual sea el objetivo que se pretende sacar del análisis, los procedimientos a seguir variarán. A pesar de las posibles diferencias, todos los análisis deberían estar compuestos por diferentes fases que son la definición de las métricas, selección de los algoritmos y herramientas apropiados, y por último visualización de los datos.

Una vez se han llevado a cabo estos pasos, se podrán empezar a tomar decisiones basándonos en los *insights* descubiertos por los analistas de datos. Hay que tener en cuenta el objetivo que se persigue cuando hablamos del uso de *Big Data*, ya que la recolección de datos y compra de tecnología para el análisis sin un objetivo claro es una estrategia perdedora.

1.1.5 Tipos de análisis

Vamos a distinguir tres tipos de análisis diferentes, dependiendo de qué es lo que queremos obtener de los datos utilizados. El hecho de que el análisis esté enfocado de una manera o de otra no hace que las herramientas y métodos utilizados no puedan coincidir.

- **Análisis Descriptivo:** usualmente son análisis utilizados para informar sobre los datos pasados. Responde mayoritariamente a las preguntas “¿Qué ha pasado?” o “¿Qué acaba de pasar?”. Un ejemplo de este tipo de análisis podría ser el realizado por *Google Trends*, resultados de ventas en un periodo o por una campaña específica. Es muy útil cuando se quiere comprobar la eficacia de un producto o analizar diferentes KPI⁶s.
- **Análisis Predictivo:** consiste en realizar modelos para realizar predicciones. Es más complicado y se realiza mediante la extrapolación de tendencias y patrones presentes al futuro. Responde a las preguntas “¿Qué podría pasar?” y “¿Cuáles serían las razones para que ocurriese?”.
- **Análisis Prescriptivo:** es el análisis que lo que busca son soluciones, está más orientado a acciones y toma de decisiones. Responde a las preguntas “¿qué debería hacer?” y “¿Por qué debería hacerlo?”.

Los tres tipos de análisis ya se realizaban antes de que existiese el *Big Data*, aunque las herramientas que este proporciona facilitan el análisis y hace que las predicciones sean

⁶ KPI: *Key Performance Indicators*, son indicadores clave para analizar el rendimiento en el mundo empresarial.

más cercanas a la realidad y la toma de decisiones tenga mejores resultados. Para conseguir plantear las mejores estrategias, deberían complementarse los estudios con la combinación del uso de los tres análisis.

1.1.6 Aplicaciones en la actualidad

En la actualidad, todas las industrias implementan de algún modo herramientas del *Big Data*, aunque el éxito y desarrollo de los modelos analíticos no es el mismo dependiendo del sector al que hagas referencia. Los sectores que antes han implementado estas herramientas han sido el financiero, el retail y la telecomunicación, aunque se espera un gran desarrollo en sanidad en los próximos años. Algunos ejemplos son:

EDUCACIÓN

El sector de la educación tiene que ir adaptándose a las nuevas tecnologías y el uso del *Big Data* es muy útil cuando se trata del aprendizaje y la enseñanza. No solamente por un aprendizaje con mayor profundidad ya que el acceso a bases de datos y artículos facilita la realización de trabajos por parte de los alumnos. Sino también las herramientas a disposición del profesorado que permite hacer un seguimiento mas personalizado de cada alumno. Algunas herramientas de *Big Data* aplicadas al sector de la educación podrían ser el programa *Turnitin* que revisa el porcentaje de plagio que existe en un trabajo o las pruebas online que se pueden realizar y se corrigen automáticamente, lo que proporciona un feedback instantáneo para los estudiantes.

TELECOMUNICACIÓN

Destaca el uso que se hace de las técnicas de Big Data para gestionar los problemas con el *Churn*, es decir, los clientes que quieren abandonar la compañía, la tasa de cancelación. A las empresas de telecomunicación les resulta mas económico retener a un cliente que se quiere marchar de la compañía que captar nuevos clientes, por lo que se esfuerzan en obtener modelos que les permitan predecir cuando un cliente quiere marcharse de la compañía para así poder hacerle una oferta para retenerlo.

RETAIL

El uso del *Big Data* en este sector no está solamente ligado al comercio electrónico, la mejor localización para las tiendas o el estudio de tendencias, sino que se aplica en muchos otros aspectos del sector como puede ser el aprovisionamiento de las tiendas o el estudio de posicionamiento de productos dependiendo de los análisis de las compras realizadas en un supermercado.

MARKETING

Uno de los sectores que más beneficio puede sacar de las oportunidades que brinda el buen uso de los datos que tiene a su disposición. No solamente para segmentar mejor los mercados y hacer campañas más ajustadas a las necesidades de los consumidores asegurándose un buen posicionamiento, si no para analizar el éxito que estas tienen y cuales son los fallos, con el fin de no volver a cometerlos en el futuro.

ASEGURADORAS

La personalización de las pólizas de seguros dependiendo de los hábitos de los clientes es una de las tendencias dentro del sector. Gracias a los datos que se generan con IoT, las aseguradoras son capaces de realizar predicciones con los comportamientos de sus clientes y pueden crear seguros a medida, basándose en sus hábitos.

ENTIDADES FINANCIERAS

Su uso se extiende a multitud de procesos dentro del mundo financiero, desde el estudio del riesgo de las diferentes operaciones que tienen que realizar los bancos, estudios de los posibles clientes o ver patrones de blanqueamiento de capital, hasta predicciones de tendencias en los mercados y posibles crisis.

SERVICIOS

Gracias al uso de la geolocalización y el acceso a internet han surgido nuevos servicios que no solamente funcionan gracias a los datos que generan, sino que se benefician de ello como ocurre con Amazon, Uber, Spotify o Netflix. Modelos de negocio que basan sus estrategias en los datos y mejoran la calidad de sus servicios basándose en la demanda de sus productos.

SANIDAD

Las aplicaciones en este sector tienen unas previsiones muy potentes. No solamente por la posibilidad de realizar un perfil para cada paciente que pueda ser compartido por todos los médicos a los que acuda, sino por la posibilidad de implementar medidas preventivas. Estas medidas preventivas se refieren a la realización de estudios que analizan determinados síntomas antes de una enfermedad, de manera que gracias a los wearables se puedan prevenir enfermedades.

2. Machine Learning:

Machine Learning es un método de análisis de datos que se nutre del *Big Data*. Es una de las ramas de la Inteligencia Artificial, ya que su propósito es crear modelos que aprenden automáticamente. Aprendizaje en este contexto sería entendido como la generalización de conocimiento a partir de un conjunto de experiencias. Se basa en que los sistemas pueden aprender de los datos disponibles, identificar patrones y tomar decisiones sin apenas intervención humana. Además, cuando hablamos de automáticamente, nos referimos a que su manera de entender la realidad se va ajustando según llegan nuevos datos, sin necesidad de ajustes manuales.

Tradicionalmente, cuando se utilizaban herramientas de programación con el fin de obtener predicciones, se utilizaban como inputs los datos y un modelo creado por la persona que estaba diseñando el modelo.

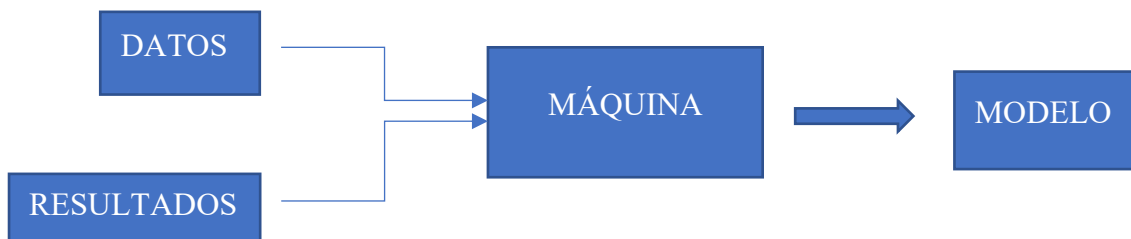
Ilustración 3 Paradigma de la programación tradicional



Fuente: Adaptación de MathWorks

Pero el uso de *Machine Learning* ha transformado por completo el procedimiento, ya que como inputs tenemos los datos y los resultados esperados, y como output queremos un modelo que poder utilizar en el futuro con otros datos nuevos.

Ilustración 4 Paradigma de Machine Learning



Fuente: Adaptación de MathWorks

Realmente lo que se hace en *Machine Learning* es crear un algoritmo que revisa los datos y es capaz de predecir comportamientos futuros, basándose en los ya experimentados.

2.1 Tipos de aprendizaje

Dependiendo de los datos disponibles para el análisis existen diferentes tipos de aprendizaje, el supervisado, en el que ya se tiene conocimiento de los resultados esperados, y el no supervisado, en el que se desconoce el resultado final, son los dos más comunes, aunque comentaré otros tipos de aprendizaje utilizados.

2.1.1 Aprendizaje supervisado

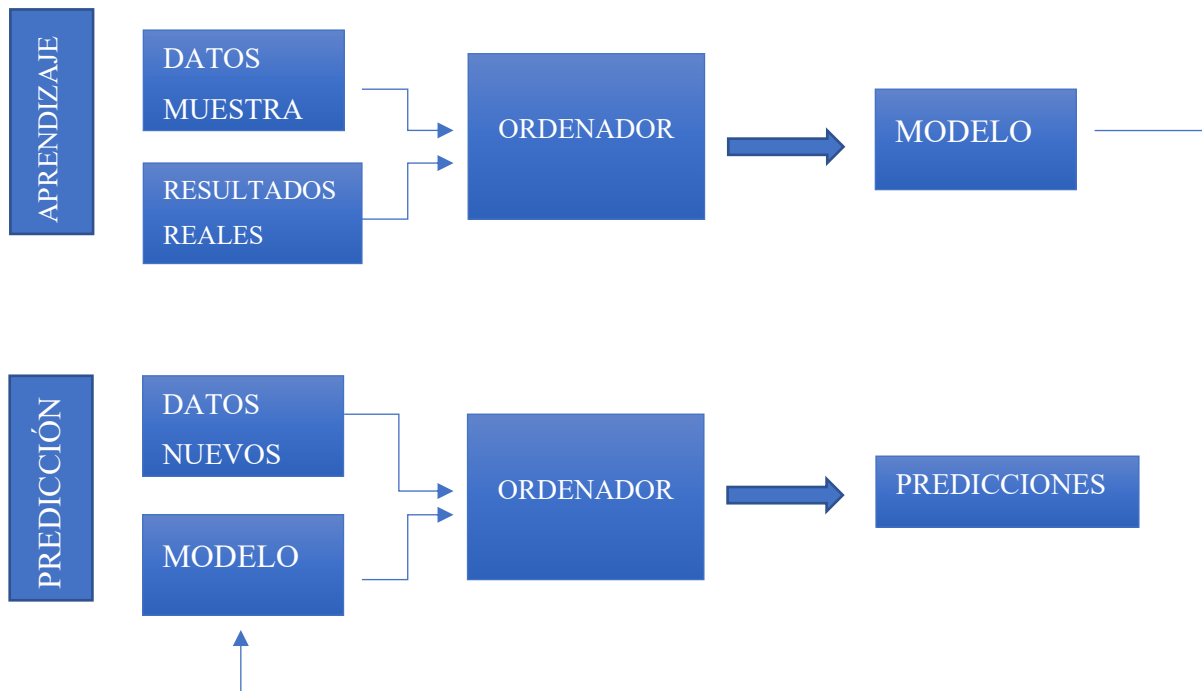
El aprendizaje supervisado crea su modelo usando datos ya clasificados, de manera que el algoritmo tiene conocimiento sobre los datos y se entrena de tal manera que su predicción coincida con el resultado ya conocido. El resultado es un modelo que predice o clasifica los datos nuevos que no han sido procesados en base al modelo creado durante la fase de aprendizaje.

Dentro del aprendizaje supervisado encontramos dos técnicas:

- **Clasificación:** se utiliza para datos con respuestas categóricas, es decir, datos que solamente tienen una respuesta de entre unos valores conocidos.
- **Regresión:** se utiliza con valores que no tienen categorías, sino que su resultado es continuo. En estos datos contamos con un número de variables explicativas, y una variable continua, que es la que nos da el resultado, de manera que trata de encontrar la relación entre las variables explicativas que más se acerque al resultado continuo.

El aprendizaje supervisado utiliza un proceso de desarrollo compuesto de dos fases, la de aprendizaje y la de predicción, independientemente de la categoría a la que pertenezcan los datos.

Ilustración 5 Proceso de desarrollo del aprendizaje supervisado



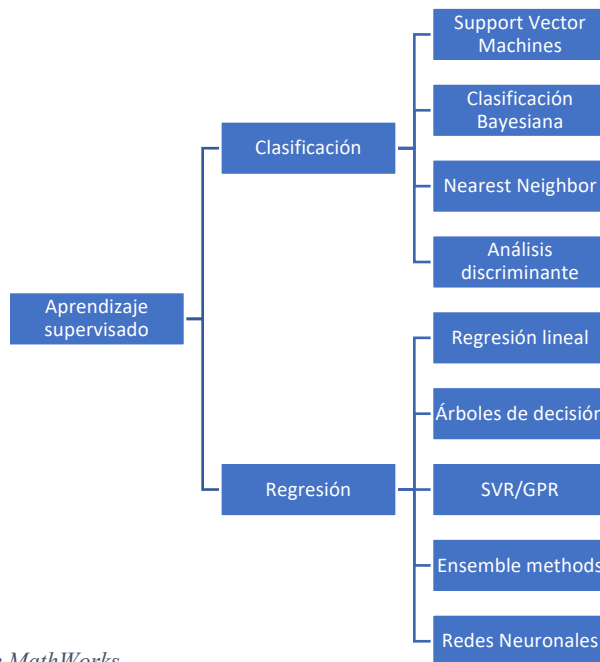
Fuente: Adaptación de MathWorks

En la primera fase del proceso, llamada de aprendizaje, se ajusta el algoritmo de manera que se crea un modelo que explique los resultados a partir de los datos de los que se parte, y se realizan los ajustes necesarios hasta que la predicción del modelo obtenido es lo más precisa posible. Una vez que el modelo conseguido sea válido, se puede utilizar para realizar predicciones con nuevos datos que no han sido procesados con anterioridad.

Como indica la definición de *Machine Learning*, aprendizaje automático de las máquinas, se busca crear un algoritmo que permita el mayor nivel de aprendizaje, por lo que los resultados predichos pueden ser metidos de nuevo en el proceso de aprendizaje junto con los datos reales con el fin de seguir afinando el modelo.

Dependiendo de cual sea el resultado que se pretende sacar del análisis, existen diferentes técnicas para conseguirlo, en la siguiente figura destaco algunos de los más importantes dependiendo de el tipo de categorías de datos en el aprendizaje supervisado.

Ilustración 6 Técnicas de análisis del aprendizaje supervisado



Fuente: Adaptación de MathWorks

2.1.2 Aprendizaje no supervisado

En el aprendizaje no supervisado los datos utilizados no están clasificados, es decir, no se tiene más información sobre los datos que los datos en si mismos. Como no se tiene conocimiento de los datos de entrada ni de los de salida, el éxito del modelo reside en la capacidad de la máquina para encontrar patrones, estructuras y relaciones dentro de los datos. La finalidad es conseguir crear grupos o reglas de asociación que permitan clasificar un dato nuevo en función de la clasificación que se ha tenido de otros datos.

Dentro de las técnicas del aprendizaje no supervisado distinguimos:

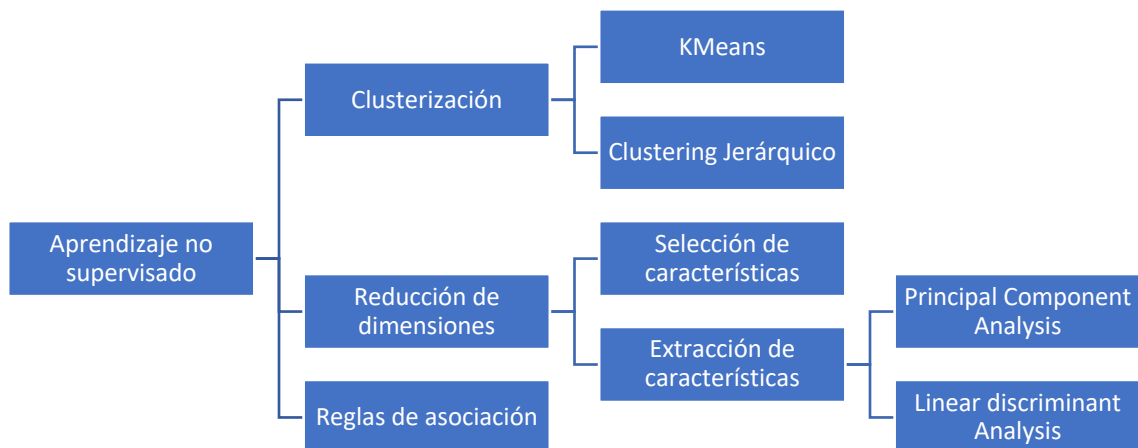
- **Clusterización:** es una técnica exploratoria que consiste en agrupar los inputs según sus características, sin necesidad de tener conocimiento de su estructura. Cada clúster (grupo) es un conjunto de objetos similares entre sí y diferentes a los otros clústers.
- **Reducción de dimensiones:** los datos vienen explicados por una alta variedad de dimensiones, las cuales muchas veces se explican unas a otras. Con este análisis

lo que se busca es reducir el número de dimensiones explicativas para realizar un análisis más sencillo a continuación a partir de las dimensiones principales.

- Reglas de asociación: busca encontrar patrones que expliquen la repetición de características presentadas simultáneamente. De manera que se pueda afirmar dentro de un nivel de confianza que si una de las características se presenta la otra también será presentada.

Dependiendo de cual sea el objetivo del análisis utilizaremos diferentes técnicas con el fin de obtener los resultados esperados. En la siguiente ilustración figuran las técnicas más comunes.

Ilustración 7 Técnicas de análisis del aprendizaje no supervisado



Fuente: Adaptación de MathWorks

2.1.3 Otros tipos de aprendizaje

Los dos tipos de aprendizaje desarrollados anteriormente son los más desarrollados dentro del mundo de *Machine Learning*, pero no son los únicos. Debido a los límites de este trabajo no se pueden desarrollar todos los tipos de aprendizaje, pero es importante comentar el aprendizaje reforzado y los algoritmos genéticos.

Este tipo de aprendizaje reforzado busca conseguir la máxima recompensa posible. Con este tipo de aprendizaje, el algoritmo se desarrolla a través de ensayo y error, y busca las acciones que produzcan mayores recompensas. Estas recompensas se basan en unos valores asociados a diferentes acciones, por lo que deberá conseguir el mayor beneficio posible, teniendo en cuenta que cada “movimiento” cuenta, por lo que además deberá ser lo más eficiente posible.

El aprendizaje genético fue introducido en los años setenta, y se inspiran en la evolución biológica como estrategia para resolver sobre todo problemas de optimización. Su manera de funcionamiento intenta replicar la selección natural y la genética, de manera que busca la mejor solución comparando un conjunto de soluciones. Además, con el fin de seguir mejorando las soluciones, se cruzan las mejores soluciones para ver si la combinación de ambas genera una solución mejor, como ocurriría en la naturaleza con la evolución.

2.2 Deep Learning y redes neuronales

Warren McCulloch y Walter Pitts, fueron los primeros en proponer en 1943 desde un marco teórico el concepto de la red neuronal, aunque no fue hasta 1949 cuando Donald Hebb explica los procesos de aprendizaje basándose en estas redes, lo que formará las bases de la *Teoría de las redes neuronales artificiales*.

Estas redes artificiales buscan asemejarse lo máximo posible a las redes neuronales reales, por lo que presentan características de aprendizaje similares, lo que ofrece numerosas ventajas como pueden ser:

- **Aprendizaje adaptativo:** se basan en aprender a llevar a cabo un aprendizaje basado en un entrenamiento, por lo que les permite adaptarse mejor. Son sistemas dinámicos auto-adaptativos, que pueden seguir aprendiendo durante su funcionamiento. El diseñador únicamente tiene que encontrar la arquitectura adecuada.

- Auto-organización: es capaz de crear su propia representación de la información que recibe mediante su proceso de aprendizaje, lo que permite su generalización, es decir, que es capaz de utilizar datos con los que no había trabajado anteriormente. También permite trabajar con datos incompletos o con ruido.
- Tolerancia a los fallos: la destrucción de parte de la estructura no impide el funcionamiento en conjunto de esta, aunque sí que deteriora los resultados obtenidos.
- Operación en tiempo real: lo que le permite adaptarse rápidamente a los cambios del medio en el que actúa.

Existen dos fases dentro de la modelización de las redes neuronales:

1. Fase de entrenamiento:

En esta fase se utiliza el conjunto de datos que se ha determinado para realizar el entrenamiento y se realizan los cálculos de los pesos que definen el modelo. Este proceso se repite hasta que se consigue un resultado lo más aproximado posible a los resultados deseados.

2. Fase de prueba:

Debido a que la fase previa ha sido realizada con unos datos determinados, es muy posible que exista un sobreajuste, por lo que se usa un segundo grupo de datos, denominados de validación, para permitir un ajuste en el modelo conseguido con el fin de poder extrapolar los resultados.

Los pesos óptimos se consiguen mediante el uso de algoritmos que permitan el mejor resultado con la menor energía posible. Dependiendo del tipo de aprendizaje, las reglas de aprendizaje, la arquitectura de la red y sobre todo la solución que se busque y la tarea para conseguirlo, se escogerá un algoritmo de aprendizaje determinado. No todos los algoritmos de aprendizaje son válidos para cualquier tipo de arquitectura a pesar de que busquen realizar la misma tarea.

Tabla 2 Algoritmos de aprendizaje más importantes

Aprendizaje	Regla de aprendizaje	Arquitectura	Algoritmo de aprendizaje	Tareas
SUPERVISADO	Corrección del error	Perceptrón o perceptrón multicapa	Algoritmos de aprendizaje, retropropagación del error, ADALINE, MADALINE	Clasificación de patrones, aproximación de funciones, predicción, control, etc.
		Elman y Jordan recurrentes	Retropropagación del error	Síntesis de series temporales
	Boltzmann	Recurrente	Algoritmo de aprendizaje de Boltzmann	Clasificación de patrones
	Competitivo	Competitivo	LVQ	Categorización intraclase, comprensión de datos
		Red ART	ARTMap	Clasificación de patrones, categorización intraclase
NO SUPERVISADO	Corrección del error	Red de Hopfield	Aprendizaje de memoria asociativa	Memoria asociativa
		Multicapa sin realimentación	Proyección de Sannon	Análisis de datos
	Competitiva	Competitiva	VQ	Categorización, compresión de datos
		SOM	Kohonen SOM	Categorización, análisis de datos
		Redes ART	ART1, ART2	Categorización
REFORZADO	Hebbian	Multicapa sin realimentación	Análisis lineal de discriminante	Análisis de datos, clasificación de patrones
		Sin realimentación o competitiva	Análisis de componentes principales	Análisis de datos, compresión de datos

Fuente: Adaptación de material académico de Universidad Carlos III

Las conexiones de pesos que existen entre dos capas diferentes pueden ser de diferentes tipos:

- Feedforward: los valores de las capas inferiores pasan hacia las capas superiores mediante conexiones hacia delante.
- Feedback: los valores de capas superiores retroceden hacia las capas inferiores.
- Lateral: en algunas redes neuronales se trabaja de tal forma que la neurona de salida que tenga el valor más alto se le asigna el valor total, y al resto cero.
- Retardo: utilizados en modelos dinámicos y temporales, donde es necesaria memoria.

El tamaño de la arquitectura de las redes depende de la cantidad de neuronas con las que cuenta cada capa y la cantidad de capas de la red. El número de capas ocultas está directamente relacionado con las capacidades de la red, aunque siempre hay que optimizar, ya que en ocasiones mayor número de capas no proporciona el comportamiento óptimo de la red.

Deep Learning hace referencia a el uso de redes neuronales de múltiples capas, en las que no hace falta supervisión, ya que la forma de aprendizaje se lleva a cabo por la propia máquina. Se proponen modelos para que realicen tareas similares a las de aprendizaje los seres humanos, como reconocimiento del habla, de imágenes y realizar predicciones a partir de unos parámetros básicos.

2.2.1 Perceptrón multicapa

El Perceptrón multicapa, también conocido como redes neuronales, se basa en la concatenación de diferentes perceptrones con el fin de obtener un resultado determinado. Su forma está inspirada en las redes neuronales del ser humano, de manera que unas neuronas o perceptrones se conectan con otras y van pasando información hasta conseguir el resultado esperado.

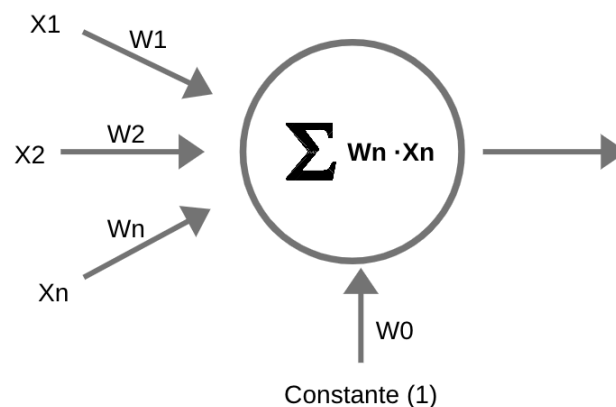
La forma de funcionamiento de un perceptrón multicapa es sencilla. Se trata de una sucesión de perceptrones simples conectados en diferentes niveles, que dan como resultado una solución basándose en los cálculos internos de cada perceptrón. Para comprender mejor su funcionamiento explicaré qué es un perceptrón y cómo funciona, cómo son las redes neuronales y algunos ajustes necesarios para su funcionamiento.

PERCEPTRÓN SIMPLE

Un perceptrón simple, también conocido como una neurona, es la unidad básica que formará una red neuronal, aunque puede funcionar de manera independiente. El diseño está inspirado en el de una neurona real, de manera que habría un número de dendritas, que son las que recogen la información, un cuerpo que es dónde se realizan las diferentes operaciones matemáticas que asignan peso a la información recogida anteriormente, y por último está el axón, que da la información resultado o pasa la información a la siguiente neurona.

Su funcionamiento es sencillo, consiste en el sumatorio de unos inputs, a los que asignará unos pesos de importancia para conseguir un resultado determinado, definiendo con qué intensidad cada variable de entrada afecta a la neurona. Además, se añade un sesgo en el que la variable siempre tiene valor uno.

Ilustración 8 Perceptrón simple



Fuente: Elaboración propia

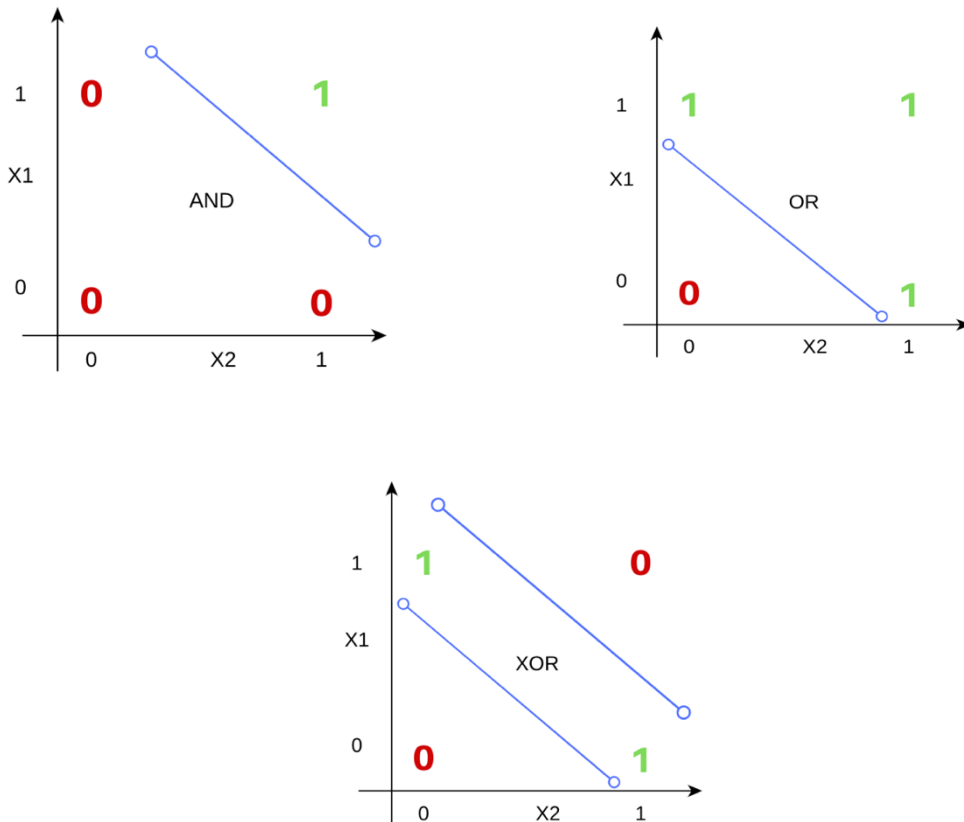
De manera que, si se deja la forma más simple de un perceptrón, tendríamos una regresión lineal, representada matemáticamente por:

$$y = \sum (w_n \cdot x_n) + w_0 \cdot 1$$

En la que podemos manipular el resultado en conjunto cambiando el peso del parámetro de la variable independiente o constante, de manera que pueda dar resultados positivos o negativos si lo que buscamos son respuestas binarias o nos ayuden a hacer una aproximación más real en caso de contar con un resultado continuo.

Dependiendo de la cantidad de variables, hablaremos de un plano bidimensional o de planos más complejos. Una neurona sola puede resolver problemas de clasificación de tipo “and” y “or”, ya que son fácilmente diferenciables con una recta, pero cuando son problemas más complicados como los de tipo “xor”, hace falta usar una segunda neurona para poder resolver el problema.

Ilustración 9 Gráficas de problemas resueltos mediante funciones a partir de un perceptrón simple o varios perceptrones simples

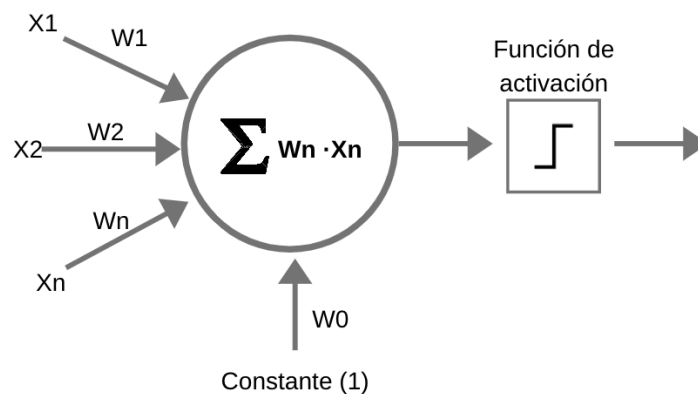


Fuente: Elaboración propia

Dependiendo de la complejidad de los datos utilizados y de las variables que se quieran utilizar, hará falta un número cada vez mayor de perceptrones, por lo que se construyen redes neuronales.

El problema se encuentra en que la combinación de regresiones lineales hace una nueva regresión lineal, y hay problemas que no se pueden resolver con una regresión lineal, por lo que se añade a cada neurona una función de activación.

Ilustración 10 Perceptrón simple con función de activación



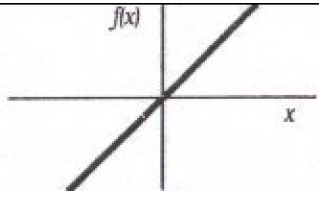
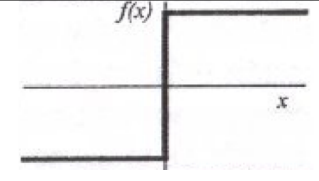
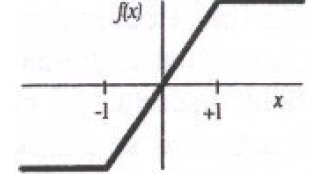
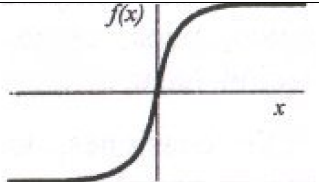
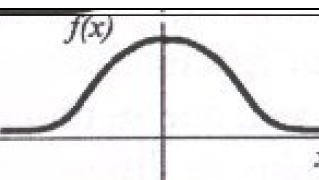
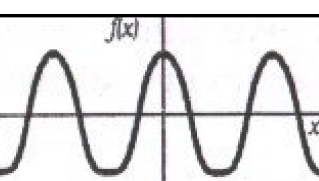
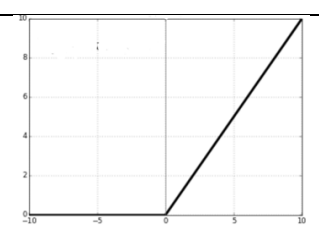
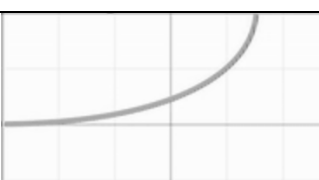
Fuente: Elaboración propia

Una función de activación se utiliza para darle la forma deseada al resultado de los cálculos realizados en una neurona, de manera que no tenga forma de regresión lineal, y el resultado final de la combinación de neuronas quede con un resultado diferente a una regresión lineal.

A continuación, explico en una tabla las principales funciones de activación más utilizadas en las redes neuronales. La tabla está compuesta por el nombre de la función, la ecuación matemática que la representa, el rango de valores salida y una gráfica bidimensional de la función.

Hay que tener en cuenta que por lo general se incluyen más de dos variables, por lo que las funciones y los resultados estarán representados por planos y no por rectas.

Tabla 3 Principales funciones de activación

Nombre	Función	Rango	Gráfica
Identidad	$y = x$	$(-\infty, \infty)$	
Escalón	$y = \begin{cases} 1, & \text{si } x \geq 0 \\ 0, & \text{si } x < 0 \end{cases}$ $y = \begin{cases} 1, & \text{si } x \geq 0 \\ -1, & \text{si } x < 0 \end{cases}$	$[0,1]$ $[-1,1]$	
Lineal a tramos	$y = \begin{cases} 1, & \text{si } x > 1 \\ x, & \text{si } -1 \leq x \leq 1 \\ -1, & \text{si } x < -1 \end{cases}$	$[-1,1]$	
Sigmoidea	$y = \frac{1}{1 + e^{-x}}$ $y = \tanh(x)$	$[0,1]$ $[-1,1]$	
Gaussiana	$y = Ae^{-Bx^2}$	$[0,1]$	
Sinusoidal	$y = A \text{sen}(w_x + \varphi)$	$[-1,1]$	
ReLU	$y = \begin{cases} 0, & \text{si } x < 0 \\ x, & \text{si } x \geq 0 \end{cases}$	$[0, \infty)$	
Softmax	$y = \frac{e^{x_i}}{\sum_{j=1}^J e^{x_j}}$	$(0,1)$	

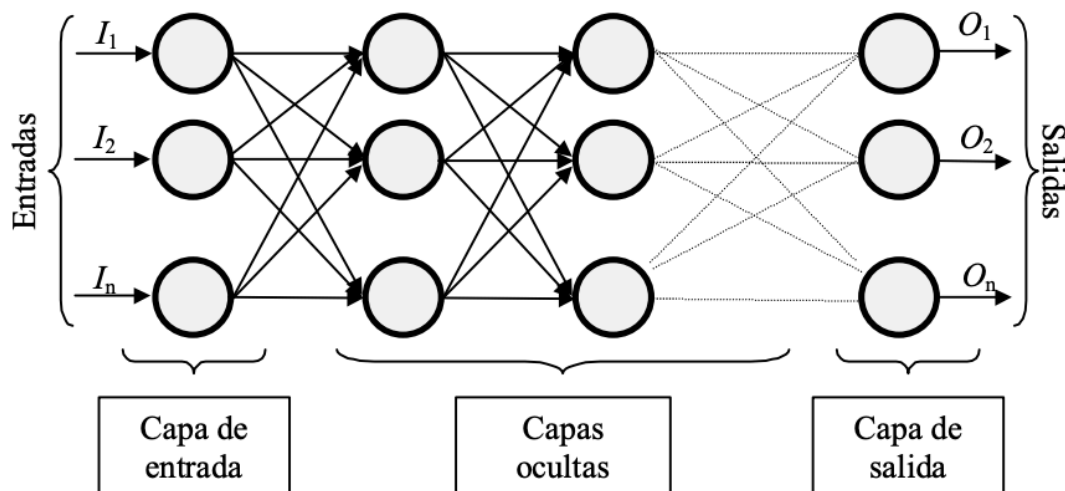
Fuente: Adaptación de Barrera J.A. T. (2018)

Una vez que hemos comprendido cómo funciona un perceptrón simple y las funciones de activación podemos pasar a crear redes más complejas, cuando la red está formada por una única capa de neuronas que cumplen la función de neuronas de entrada y de salida simultáneamente se denomina red monocapa, mientras que si contamos con dos o más capas se denominan multicapa.

PERCEPTRÓN MULTICAPA

Como ya he mencionado anteriormente, un perceptrón multicapa está formado por la concatenación de diferentes perceptrones. Estos perceptrones se pueden organizar o bien en la misma capa, o formando diferentes capas, y los resultados de la primera capa serán los datos de entrada de las siguientes capas, hasta llegar a la solución esperada.

Ilustración 11 Esquema de un perceptrón multicapa (MLP)



Fuente: Matich, D. J. (2001).

Todos los perceptrones multicapa están compuestos por una capa de entrada, una capa de salida y n capas ocultas entre la entrada y la salida, que es donde se realizarán la mayoría de los ajustes automáticamente. Cada neurona tiene salidas distintas, pero están relacionadas entre sí, ya que la salida de una neurona es la entrada de la siguiente, lo que permite un aprendizaje jerarquizado.

La manera de entrenar el modelo está, como ya mencioné anteriormente, dividido en dos fases, una llamada de aprendizaje en la que la red ajusta sus pesos con el fin de conseguir un resultado lo más similar posible al resultado esperado y la de validación, en la que con otro conjunto de datos termina de entrenar a la red con el fin de erradicar el sobreentrenamiento.

Para ello se realiza una prueba de *cross validation*, con el fin de disminuir lo máximo posible este sobreentrenamiento de la red y conseguir un modelo robusto capaz de trabajar con datos que no ha observado nunca. Esta prueba consiste en la separación de los datos en diferentes grupos, de manera que primero se utiliza un grupo de datos para entrenamiento y luego se comprueba el resultado antes de ponerlo en marcha, y se realizan los ajustes necesarios para que su funcionamiento sea óptimo con los datos nuevos que puedan ser utilizados.

La separación de los datos en diferentes grupos para realizar esta prueba es algo complicado, ya que dependiendo de la naturaleza de los datos serán necesarios diferentes separaciones. Cuando hablamos de predicciones de series temporales la separación que se realiza es de diferentes grupos teniendo en cuenta el tiempo y no de manera aleatoria, ya que no puedes entrelazar la línea temporal, si lo que se busca es resolver problemas de clasificación, es posible realizar una separación aleatoria.

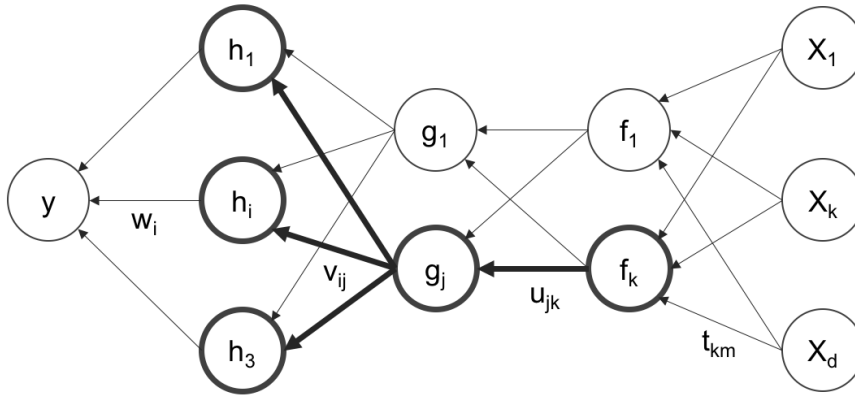
La creación de los diferentes grupos de datos para hacer *cross validation*, depende estrictamente de los datos que vayan a ser analizados y es importante tenerlo en cuenta.

Una vez se ha comprobado que existe sobreentrenamiento por parte del modelo se pueden usar diferentes herramientas para resolverlo, una de ellas es el *back-propagation*, se utiliza una vez se ha terminado el modelo inicial.

Lo que pretende este algoritmo de aprendizaje es encontrar el punto donde el peso no es el adecuado para generalizar el modelo, para ello en lugar de seguir el mismo recorrido que se ha usado para hacer las estimaciones, parte desde el punto final de la red,

calculando cual es el error entre lo esperado y la predicción, y realiza ajustes en el perceptrón que desajusta la red.

Ilustración 12 Representación de la función de Back-propagation en un perceptrón multicapa

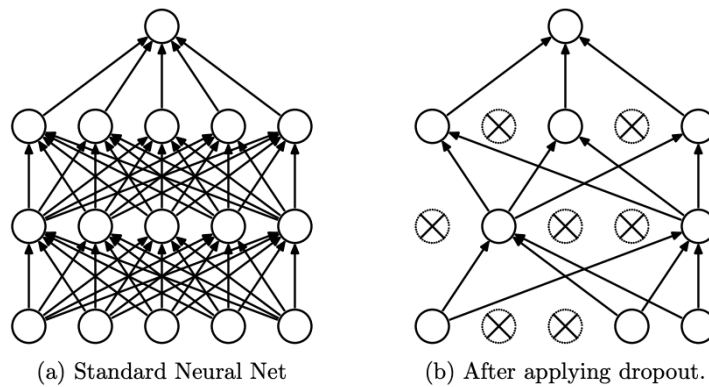


Fuente: Rajalingappaa Shanmugamani (2018)

Debido a que el aprendizaje de las capas es jerarquizado, solo hay que corregir aquellas neuronas que estén conectadas a la neurona en la que se encuentre el fallo, y no todas las neuronas, por lo que además es más sencillo.

Por último, es importante tener en cuenta el *dropout*, que también sirve para evitar el sobre entrenamiento. Funciona de manera que algunas de las neuronas se eliminan dependiendo de la información que generen puesto que están muy especializadas en algunos datos y los resultados que dan no son generalizables.

Ilustración 13 Representación de la función dropout en un perceptrón multicapa



Fuente: Hochreiter, S., & Schmidhuber, J. (1997)

El mayor problema al que se enfrentan las redes neuronales es al sobre entrenamiento, ya que en ocasiones se hiper-especializan en los datos de entrada que han recibido y luego no se obtienen resultados con datos nuevos, ya que lo que ha hecho ha sido “memorizar” las características de los datos de entrada iniciales.

Este tipo de redes neuronales se utilizan para resolver sobre todo dos tipos de problemas:

- Problemas de predicción: estimación de una variable continua a partir de unas variables de entrada (continuas o discretas).
- Problemas de clasificación: asignación de una categoría basándose en las características de las variables de entrada y patrones.

2.2.2 Redes recurrentes

Las redes recurrentes son un tipo de red neuronal en el que se incluyen bucles, pueden ser de una neurona consigo misma, entre diferentes neuronas de una misma capa o entre neuronas de diferentes capas, lo que permite un aprendizaje más profundo y complejo. Para el funcionamiento de estas redes es necesario añadir la variable tiempo, de manera que su activación queda definida por:

$$a_i(t+1) = f\left(\sum_j w_{ji} \times a_j(t)\right)$$

Siendo a_j la función de activación de todas las neuronas que poseen conexión con la neurona i . La variable tiempo hace que estas redes tengan un comportamiento dinámico y temporal, lo que es clave en la predicción de secuencias temporales. La evolución de estas activaciones es continua, ya que dependen del momento temporal anterior, por lo que en este caso el aprendizaje será continuo o por épocas.

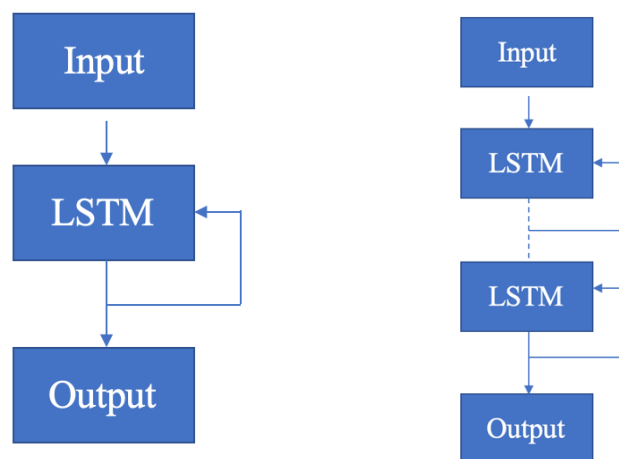
Estas redes pueden ser parcialmente recurrentes, en las que solo determinadas neuronas o capas cuentan con esta manera de procesar el tiempo, o totalmente recurrentes, pero todas utilizan algoritmos supervisados para el ajuste de parámetros.

Existen muchas redes recurrentes como la de Hopfield que es un modelo de memoria asociativa y se utiliza sobre todo para recuperar datos incompletos a partir de asociación mediante conexiones entre todas sus neuronas, algunas parcialmente recurrentes como la red de Jordan que utiliza neuronas recurrentes en la capa de entrada, y la red de Elman en la que las neuronas de la capa de entrada reciben la información de las capas ocultas, o las LSTM, que explicaré más en profundidad, ya que son muy útiles para las predicciones de series temporales.

REDES LSTM

Las redes *Long short-term memory* (LSTM) presentan una arquitectura de redes propuesta por Hochreiter y Schmidhuber, (1997), se compone por unidades dentro de la capa oculta a la que denominan *memory blocks*. Cada uno de estos bloques puede contener varias células y una red puede contener más de un bloque.

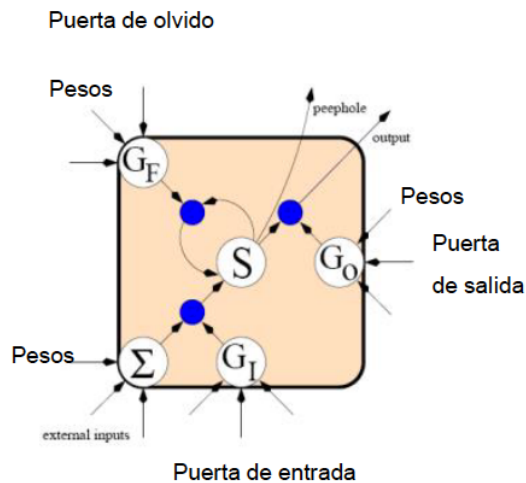
Ilustración 14 Esquema de redes LSTM con una o varias memory blocks



Fuente: Adaptación de material académico de Universidad Carlos III

El interior de cada uno de estos bloques es complejo, dentro pueden contener diferentes números de neuronas, junto con las tres puertas que controlan la activación de la célula, que son la de entrada, la de olvido y la de salida.

Ilustración 15 Esquema de un memory block y sus componentes principales



Fuente: Adaptación de material académico de Universidad Carlos III

Dentro de cada *memory block* es interesante destacar diferentes procedimientos que las hace más efectivas ante el análisis de secuencias temporales.

Uno de ellos es la capacidad de memorizar, por lo que puede hacer frente a series temporales muy extensas. Además, cuenta con una puerta de olvido por lo que puede decidir olvidar parte de la información que recibe y no toda como ocurre con el dropout. Esta memoria parcial le permite mejorar sus predicciones, ya que se basa solo en la información que considera realmente relevante. Por último, destacaré que cuenta con una reiteración que pasa por el punto de olvido, lo que le permite analizar la información una vez que ha sido utilizada, para afinar más los datos que realmente son necesarios.

El objetivo de su entrenamiento consiste en minimizar el error de las predicciones que se obtienen de la red. Para ello utiliza las herramientas de la retropropagación y de descenso del gradiente, aunque el aprendizaje de este tipo de redes es complejo, por lo que se utilizan herramientas de Deep Learning y de computación evolutiva.

3. Machine Learning en Trading:

Como ya hemos visto, las aplicaciones que se pueden dar al *Machine Learning* son innumerables en diversos campos, pero yo me voy a centrar en las aplicaciones en trading. El trading hace referencia a las transacciones de compraventa de activos en mercados financieros, generalmente a corto y medio plazo, con el fin de obtener una plusvalía. Los agentes que operan en estos mercados utilizan una gran variedad de herramientas con el fin de predecir los movimientos del mercado y los precios de los activos financieros.

Gracias a las facilidades de cálculo que nos ofrece el *Machine Learning*, se pueden hacer mejores predicciones en trading, aunque no es el único uso que tiene el *Machine Learning* dentro del mundo del trading, ya que también se puede utilizar para calcular el riesgo de los activos o las fronteras eficientes de inversión en cada mercado.

Para poder comprender mejor el funcionamiento del trading hay que tener unas nociones básicas sobre los mercados financieros, sus activos y los posibles riesgos que conlleva el realizar operaciones en ellos.

Lo primero que es importante tener claro es los tipos de mercado que existen y la manera en la que se opera en cada uno de ellos. Existen muchas formas de clasificarlos, pero debido al propósito de este estudio, vamos a centrarnos en dos, el tipo de mercado y el tipo de activo de cada mercado.

Distinguimos principalmente tres tipos de mercados:

- Mercado primario: donde se emiten por primera vez los activos financieros.
- Mercado secundario: donde se realiza en intercambio de títulos valores previamente emitidos, como por ejemplo el mercado de acciones (la Bolsa).
- OTC: por sus siglas en inglés *Over the Counter*, es como se conoce a los mercados no organizados, donde se realizan transacciones directamente entre las dos partes interesadas. Generalmente son operaciones a medida por necesidad de financiación entre entidades financieras y empresas.

Tabla 4 Tipos de activos disponibles en los Mercados financieros

Renta fija y Renta variable <ul style="list-style-type: none">• Acciones y Bonos• Private Equity
Commodities <ul style="list-style-type: none">• Materia primas (trigo, café, maíz, etc.)• Sector Inmobiliario• Petróleo• Minerales
Divisas <ul style="list-style-type: none">• Moneda extranjera y tipos de cambio• Criptomonedas

Fuente: Elaboración propia

En estos activos se puede invertir de manera directa, a través de una sociedad de gestión colectiva, como son los fondos de inversión (*Hedge Funds*), ETFs (fondos de inversión cotizados) o sociedades de inversión de capital variable, y mediante el uso de derivados.

Otros términos importantes para tener en cuenta son la rentabilidad, la volatilidad y el riesgo. Estos son los que determinarán la inversión y los beneficios que pueda proporcionar al inversor.

La rentabilidad hace referencia a los beneficios que se han obtenido o se pueden obtener de una inversión. El objetivo es siempre obtener la mayor rentabilidad posible, aunque teniendo en cuenta el riesgo que haya que asumir para conseguirlo. Se denota con la letra griega Delta (Δ), matemáticamente se entiende como los cambios en el precio del activo.

La volatilidad hace referencia a la frecuencia e intensidad de los cambios en el precio de un activo financiero. La volatilidad de un activo en relación con el mercado se denomina por el coeficiente beta (β).

Por último, está el riesgo, que hace referencia a la incertidumbre sobre la obtención de un resultado determinado. Hay diferentes tipos de riesgo posibles y diferentes clasificaciones según el riesgo en trading.

De entre los diferentes tipos de riesgo que afectan a los mercados financieros, riesgo de crédito o impago, riesgo de inflación, riesgo de liquidez, riesgo de tipo de cambio, riesgo operacional o riesgo de la volatilidad, hay que preocuparse sobre todo por dos. El más importante es el riesgo de crédito, que es el que dará calificación a los distintos activos financieros, y también habría que tener en cuenta el operacional, ya que al ser operaciones llevada a cabo por personas existe la posibilidad de errores. El resto de los riesgos es posible solventarlos mediante diferentes métodos como son cubrir las posiciones o la diversificación.

Los ratings de los distintos activos financieros publicados por las agencias de calificación de riesgo indican según el criterio de cada agencia el riesgo de crédito, es decir la posibilidad de impago de los diferentes activos. Las agencias de crédito más destacadas son *Standard & Poor's*, *Moody's*, *Fitch*, y el indicador de mercados europeo (EIU), y cada una tiene una notación diferente para dar su calificación según el riesgo.

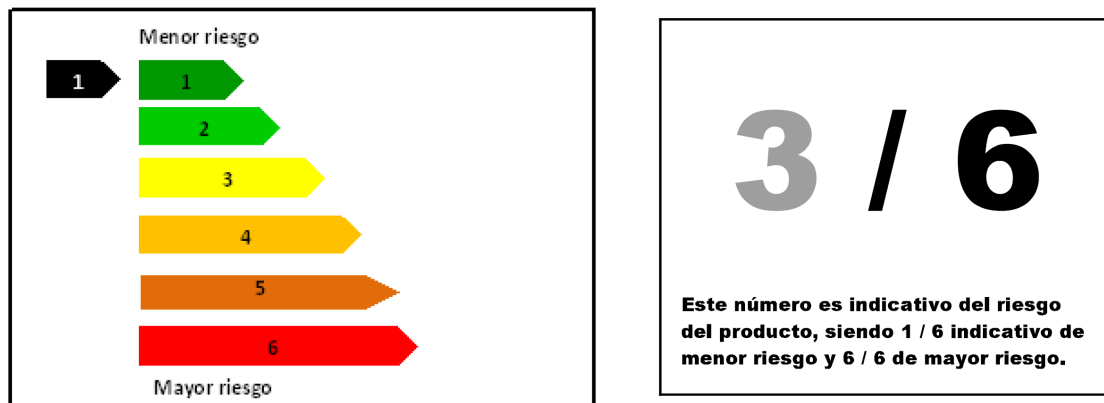
Tabla 5 Ratings de los activos financieros en función de diferentes agencias de rating

EUI Market Indicators	Fitch/ S&P's		Moody's	Rating grade description (Moody's)	
A(1)	AAA		Aaa	Nivel de inversión	Riesgo de crédito mínimo
	AA+		Aa1		Riesgo de crédito muy bajo
	AA		Aa2		
AA-		Aa3			
B(2)	A+		A1		Riesgo de crédito bajo
	A		A2		
	A-		A3		
	BBB+		Baa1	Riesgo de crédito moderado	
	BBB		Baa2		
	BBB-		Baa3		
C(3)	BB+		Ba1	Nivel de Especulación	Riesgo de crédito sustancial
	BB		Ba2		
	BB-		Ba3		
	B+		B1		Riesgo de crédito alto
	B		B2		
	B-		B3		
D(4)	CCC+		Caa1	Riesgo de crédito muy alto.	
	CCC		Caa2		
	CCC-		Caa3		
	CC		Ca	Cerca de quiebra, posibilidad de recuperación	
	C				
E(5)	DDD	SD	C	Quiebra, no existe recuperación de la inversión	
	DD	D			
	D				

Fuente: Adaptación de The Economist Data Tool

En España, entró en vigor en 2016 una nueva forma de clasificar y dar información sobre los diferentes activos financieros. Esta manera de clasificar los activos financieros busca mayor simplificación de la información, para hacerla accesible y comprensible para las personas que no tengan un amplio conocimiento de los mercados financieros.

Ilustración 16 Niveles de riesgo de activos financieros según la legislación española



Fuente: BOE <https://www.boe.es/buscar/act.php?id=BOE-A-2015-11932>

Dependiendo de la fase de inversión en la que te encuentres, de los activos con los que se trabaja y la teoría de mercado que utilice el agente, las herramientas utilizadas serán diferentes. De manera que, para facilitar un poco la comprensión de las diferentes herramientas utilizadas, dividiremos las posibles estrategias por el tiempo en el que se mantiene la posición de la inversión.

HIGH FREQUENCY TRADING

Esta manera de operar en los mercados financieros es relativamente nueva, ya que se puede realizar gracias a los intercambios electrónicos y la rapidez de la transmisión de datos. Los horizontes de inversión son menores a un día, generalmente se trata de minutos el tiempo que se mantiene una posición, y al final del día se cierran todas las posiciones, independientemente de cual sea su valor. Generalmente utiliza el mercado de valores, comúnmente conocido como la Bolsa.

Estas estrategias están basadas en los diferentes algoritmos de análisis de los que se dispone y la rapidez en la que se puedan realizar las transacciones. Por lo tanto, el éxito

de las operaciones reside en la capacidad analítica y de decisión que tenga el programa que se utilice. La manera de funcionamiento de estos mecanismos suele ser muy similar, sin tener en cuenta la herramienta o los algoritmos utilizados.

Los datos utilizados son los *ticks*, que indican los precios en un momento determinado. Difieren dependiendo de la frecuencia de tiempo que se elija, generalmente se utilizan intervalos de un minuto a cinco minutos en este tipo de trading.

Otra característica que tienen en cuenta estos modelos es la horquilla de precio en la que operan, conocido como el "*bid-ask spread*". Esto hace referencia al precio que el comprador está dispuesto a pagar y el que el vendedor está dispuesto a aceptar. Si los precios están dentro de la horquilla no se tienen que realizar movimientos, cuando el precio está por debajo del mínimo se recomienda comprar y cuando sube por encima del máximo vender.

Por último, tienen en cuenta el *Mean reversion*, una teoría del trading, que postula que el precio tiende a estabilizarse en su precio medio a largo plazo, por lo que de los pequeños desajustes es de donde sacan provecho este tipo de transacciones.

Una de las herramientas más utilizadas en este tipo de transacción son las redes neuronales con múltiples capas. En su estudio, P. Ganesh y P. Rakheja (2018) proponen construir un modelo MLP (perceptrón multicapa) que estudie las tres posibilidades de precio; un incremento en el precio, un descenso en el precio o estabilidad del precio, con un nivel de confianza para cada clase. En caso de que la predicción no supere el nivel de confianza mínimo no se procedería a realizar ningún tipo de transacción. Además, proponen un modelo que va restableciendo los parámetros según va actualizando la información del mercado, ya que se observa que sin esta actualización periódica la precisión es menor.

Como he mencionado anteriormente, cuando se trabaja con redes neuronales hay que realizar ajustes con las diferentes funciones de activación, en este caso han utilizado la función *ReLU* en todas las capas excepto en las dos últimas que utilizan capas *Softmax*

para poder comparar todos los modelos posibles y escoger el más aproximado a la realidad. También se realizan estudios para ver cual es el nivel óptimo de neuronas en cada capa, ya que en ocasiones no por tener más capas o perceptrones se llegan a mejores resultados, se busca optimizar la red y se hace mediante un estudio de la optimización, comparando diferentes modelos terminados.

Pero no se utilizan solamente las redes neuronales en este tipo de trading, también se han realizado estudios utilizando aprendizaje reforzado, ya que el objetivo es conseguir la maximización de beneficios con las transacciones o el uso de herramientas más simples, que ofrecen resultados buenos en relación con las inversiones necesarias para ponerlas en funcionamiento, como muestra el estudio de Kearns y Nevmyvaka (2013).

En el fondo, cualquier modelo de Machine Learning aplicado al *high frequency trading* debería dar buenos resultados, la diferencia entre ellos será la capacidad que tengan los programas utilizados para realizar el máximo análisis posible en el menor tiempo.

Por tanto, el factor más importante en estos modelos será la disposición de herramientas y capital de quien construya el algoritmo, ya que será más fácil crear redes más complejas y mejor entrenadas para su uso.

TRADING A CORTO PLAZO

Es el tipo de trading más común, son transacciones especulativas que se hacen teniendo en cuenta las previsiones del día siguiente con el fin de conseguir beneficios. No se hacen inversiones por interés en una empresa o los posibles dividendos que puedan repartir anualmente, sino por las posibilidades de obtener una plusvalía.

El uso de técnicas de Machine Learning es muy diverso en este campo, diferenciándose sobre todo dos aspectos, si hay un agente que utiliza las herramientas y algoritmos como parte de su análisis o si es la propia máquina la que automáticamente realiza las

operaciones, aunque sólo cambia la manera de realizar las transacciones y no el análisis previo.

Para el análisis de los mercados hay que tener en cuenta primero la eficiencia del mercado, es decir la cantidad de información y la manera de procesar la información que existe en el mercado. Este procesamiento suele estar sesgado por la percepción del analista en cuestión y sus sentimientos hacia el mercado (*bearish* o *bullish*), por lo que sus predicciones y acciones también lo estarán. Con el uso de algoritmos y *Machine Learning* ya no existe este problema, ya que sólo se basan en el análisis de los datos, creando un mercado más eficiente que el percibido por los traders.

El otro aspecto importante a la hora de realizar un análisis de este tipo es la aversión al riesgo del inversor, ya que, dependiendo de la postura elegida, el nivel de beneficios posibles será diferente. La rentabilidad de un activo está directamente relacionada con el riesgo de este, por lo que los activos con más riesgo tienen una mayor rentabilidad que los menos arriesgados.

Teniendo en cuenta de que se tratan de transacciones especulativas y que es necesario que tengan riesgo, sobre todo el asociado con la volatilidad de los activos, se pueden escoger diferentes niveles de riesgo. Los estudios académicos escogen una postura neutra ante el riesgo, con el fin de conseguir resultados más objetivos, cuyos resultados se puedan extrapolar a otros niveles de riesgo sin cambios muy bruscos en los resultados.

Con el fin de poder analizar las múltiples redes neuronales artificiales y ajustes que se realizan, voy a representar en una tabla algunos de los diferentes trabajos revisados y sus herramientas, con el fin de hacer una aproximación a la realidad de estas actividades. La elección de estos trabajos y no otros ha sido que ofrecen las mejores predicciones a corto plazo, por lo que son las herramientas en las que me voy a centrar.

Tabla 6 Herramientas que proporcionan mejores resultados en predicción de precios en mercados financieros

Autores	Descripción del trabajo realizado	Herramientas
Sharang, A., & Rao, C. (2015).	Con el fin de evitar altas volatilidades proponen un estudio con el uso de derivados. Utilizan un modelo RBM que ayuda a estimar la probabilidad de distribución del resultado.	<i>Restricted Boltzmann Machine</i>
Moghaddam, A. H., Moghaddam, M. H., & Esfandyari, M. (2016).	Para su estudio proponen un perceptrón multicapa compuesto por tres capas ocultas (20-40-20), buscando el mínimo error R^2 . Para perfeccionar su aprendizaje utilizan un algoritmo de retropropagación.	Red Neuronal alimentada con retropropagación
Navon, A., & Keller, Y. (2017).	En su trabajo, proponen una red neuronal con 5 capas, en cada una de ellas utilizan la función de activación <i>ReLU</i> . Para corregir los errores de sobre entrenamiento realizan <i>dropout</i> tras la primera y la segunda capa. Por último, para simplificar sus resultados ponen una capa de <i>Softmax</i> , con el fin de simplificar los resultados obtenidos.	<i>Deep Learning</i>
Heaton, J. B., Polson, N. G., & Witte, J. H. (2016)	Su propuesta se basa en el uso de redes recurrentes, que tienen en cuenta el paso del tiempo. Además, el uso de un modelo LSTM permite seleccionar solo la información ralmente importante, sin tener que olvidar toda la información que proporcionan determinadas neuronas.	Modelo LSTM

Fuente: Elaboración propia

Estos modelos son los que obtienen los resultados más ajustados a la realidad, con errores de entre el 5-10%, lo que permite a los analistas realizar operaciones con un riesgo menor al que tienen los activos financieros con los que se opera.

El problema con el que se encuentran todos estos modelos es el sobre entrenamiento, por lo que es necesario utilizar algoritmos que permitan evitar el sobre ajuste debido a la especialización de los datos, y estos modelos mencionados anteriormente son los que menos problemas de generalización presentan, aunque necesitan pequeños ajustes.

El más utilizado como he explicado en el apartado del Perceptrón multicapa es el uso del *dropout*, aunque elimina más información que la que proporciona un sobre ajuste, aunque estos modelos son los que menos problemas de sobre ajuste presentan.

El modelo de Boltzmann es el más simple de los utilizados, puesto que cuenta solo con una capa de variables visibles, que son los inputs, y una capa oculta a la que le asocia una ponderación determinada para explicar las variables que determinan los inputs.

Las redes LSTM son las que mejores resultados proporcionan para este tipo de predicción, sin necesidad de complicar demasiado la arquitectura del modelo. Su estructura interna permite olvidar sólo parte de la información que no es necesaria ya que conduce al sobre entrenamiento, con la ventaja de no perder información que puede resultarle valiosa. También son las más adecuadas ya que tienen en cuenta el tiempo y pueden discernir entre datos pasados y recientes, dando más importancia a los recientes, ya que es más probable que sigan esa tendencia a la que tenían el año anterior.

TRADING A MEDIO Y LARGO PLAZO

En este tipo de trading ya entraríamos en una postura más parecida a creación y optimización de carteras, que en estrategias de inversión siguiendo las tendencias especulativas a corto plazo del mercado. Se centran en los activos financieros con un mayor rating, y no en aquellos de uso especulativo.

Teniendo en cuenta que el horizonte es mucho mayor a los mencionados anteriormente hay que considerar, además de los datos históricos de los precios de las acciones, los datos macroeconómicos y las previsiones de estos, que marcarán los futuros ciclos económicos.

Para realizar el análisis de los valores de mercado se pueden tomar diferentes enfoques, como son el *top-down*, en el que escoges a partir de análisis macroeconómicos y de la industria en qué empresas invertir, o el *bottom-up*, en el que escoges un activo financiero y vas ampliando la imagen hacia el sector, y los datos macroeconómicos para así poder elegir la postura de inversión más rentable.

También es necesario considerar cual es la estrategia de gestión de esos activos que se quiere realizar, ya que, aunque el plazo sea superior al especulativo existe la posibilidad de invertir de cara a un horizonte temporal y no indefinidamente.

Estas premisas son las que ayudarán a modelar un algoritmo que se ajuste a las necesidades que tenga un inversor, que ayudado por las herramientas de *Machine Learning*, podrá construir de manera óptima su porfolio y hacer un seguimiento más exhaustivo de la evolución de los mercados una vez ha realizado la inversión.

El tener la opción de obtener información de cómo afectan los cambios en los activos financieros que se poseen permite reducir el riesgo y tomar medidas preventivas ante posibles tendencias futuras.

Para la realización de estos análisis se utilizan herramientas de *Machine Learning* más variadas. Este enfoque no se centra tanto en la predicción de los precios como en la optimización de los activos disponibles en base a unas premisas determinadas por el inversor, por lo que se utilizarán herramientas enfocadas a la clasificación de activos financieros. Debido a su sencillez y buenos resultados es muy útil utilizar modelos de Clasificación Bayesiana⁷, así como de agrupación de activos según sus características como ocurre con el modelo de *Matchbox*, un recomendador híbrido que combina el análisis de contenido con un filtrador colaborativo que comprueba si los resultados obtenidos se ajustan a la postura del inversor.

⁷ Clasificador Bayesiano: basado en el Teorema de la probabilidad condicionada, son modelos que predicen la probabilidad de posibles resultados.

4. Conclusiones

Con el auge del *Big Data* las herramientas de *Machine Learning* han ido tomando cada vez más protagonismo en el mundo empresarial. El trading es una actividad que desde que surgió ha estado actualizado con las herramientas de análisis de los mercados y la posibilidad de crear predicciones cada vez más ajustadas a la realidad, y el uso de las herramientas de *Machine Learning* le permite acercarse más a esas predicciones.

Con este trabajo se pretendía hacer una labor de investigación que ayudase a comprender y sintetizar cuales son las herramientas de *Machine Learning* más utilizadas hoy en día y por qué son las herramientas más adecuadas para cada tipo de trading, adaptando las características necesarias para una buena comprensión de los mercados actuales.

Como resultado se ha conseguido realizar un análisis de las herramientas básicas y los algoritmos de aprendizaje necesarios para conseguir resultados extrapolables a otros datos que todavía no son conocidos, ayudando así a la toma de decisiones de los inversores. Se ha conseguido acotar una base sobre la cual se puedan realizar estudios más en profundidad con el fin de crear aplicaciones reales de estas herramientas en el campo del trading teniendo en cuenta los horizontes de inversión.

El crecimiento en la generación de datos es exponencial, y no se prevé que se vaya a estabilizar en los próximos años, por lo que el uso de herramientas como *Machine Learning* también promete un alto desarrollo. Una de las herramientas que más desarrollo está consiguiendo es la de *Text mining*, lo que posibilita acceder a información primaria sobre el mundo financiero y como lo perciben los inversores. Este acceso a información primaria es la clave para el desarrollo de análisis de los mercados financieros teniendo en cuenta el sentimiento de los inversores y cómo estos van a invertir su dinero, lo que en última instancia es el verdadero motor de los mercados financieros.

5. Bibliografía

Atsalakis, G. S., & Valavanis, K. P. (2009). Forecasting stock market short-term trends using a neuro-fuzzy based methodology. *Expert systems with Applications*, 36(7), 10696-10707.

Barrera, J. A. T. (2018) Redes Neuronales. *Universidad de Guadalajara Disponible en: http://www.cucei.udg.mx/sites/default/files/pdf/toral_barrera_jamie_areli.pdf [Visitada en mayo de 2019].*

Bechara, J. E. A., Cruz, J. C. T., & Ceballos, H. V. (2009). Predicciones de modelos econométricos y redes neuronales: el caso de la acción de SURAMINV. *Semestre Económico*, 12(25), 95-109.

Benedetti, F., Baigún, A., Yanguas, M., Heymann, D., Perazzo, R., & Zimmermann, M. (2008). Predicciones de Series de Tiempo con Redes Neuronales: Trading usando el Índice Merval.

Bhadani, A. K., & Jothimani, D. (2016). Big data: challenges, opportunities, and realities. In *Effective Big Data management and opportunities for implementation* (pp. 1-24). IGI Global.

Brewer, E. (2012). Pushing the cap: Strategies for consistency and availability. *Computer*, 45(2), 23-29.

Dash, R., & Dash, P. K. (2016). A hybrid stock trading framework integrating technical analysis with machine learning techniques. *The Journal of Finance and Data Science*, 2(1), 42-57.

De Mauro, A., Greco, M., & Grimaldi, M. (2015, February). What is big data? A consensual definition and a review of key research topics. In *AIP conference proceedings* (Vol. 1644, No. 1, pp. 97-104). AIP.

Decker, K., Sycara, K., & Zeng, D. (1996). Designing a multi-agent portfolio management system. In *Proceedings of the AAAI Workshop on Internet Information Systems* (Vol. 60).

El Economista. *Diccionario de términos financieros*. Obtenido de: <https://www.eleconomista.es/diccionario-de-economia/> [Visitada en mayo 2019]

Fu, T. C., Chung, C. P., & Chung, F. L. (2013). Adopting genetic algorithms for technical analysis and portfolio management. *Computers & Mathematics with Applications*, 66(10), 1743-1757.

Galván I.M. & Valls J.M. (2015) Open Course Redes Neuronales, Redes de Neuronas Recurrentes. Obtenido: <http://ocw.uc3m.es/ingenieria-informatica/redes-de-neuronas/transparencias/Tema6%20RedesRecurrentes.pdf/view> [Visitado en mayo 2019]

Ganesh, P., & Rakheja, P. (2018). Deep Reinforcement Learning in High Frequency Trading. *arXiv preprint arXiv:1809.01506*.

Gârleanu, N., & Pedersen, L. H. (2013). Dynamic trading with predictable returns and transaction costs. *The Journal of Finance*, 68(6), 2309-2340.

Günes Baydin, A., Pearlmutter, B. A., & Siskind, J. M. (2016). Tricks from Deep Learning. *arXiv preprint arXiv:1611.03777*.

Heaton, J. B., Polson, N. G., & Witte, J. H. (2016). Deep learning in finance. *arXiv preprint arXiv:1602.06561*.

Henrique, B. M., Sobreiro, V. A., & Kimura, H. (2018). Stock price prediction using support vector regression on daily and up to the minute prices. *The Journal of finance and data science*, 4(3), 183-201.

Hiransha, M., Gopalakrishnan, E. A., Menon, V. K., & Soman, K. P. (2018). NSE stock market prediction using deep-learning models. *Procedia computer science*, 132, 1351-1362.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8), 1735-1780.

Iyer, K. V. (2018). Big data analytics: The stakes for students, scientists & managers-a management perspective. *arXiv preprint arXiv:1803.05991*.

Jiang, Z., Xu, D., & Liang, J. (2017). A deep reinforcement learning framework for the financial portfolio management problem. *arXiv preprint arXiv:1706.10059*.

Kearns, M., & Nevmyvaka, Y. (2013). Machine learning for market microstructure and high frequency trading. *High Frequency Trading: New Realities for Traders, Markets, and Regulators*.

Laney, D. (2001). 3D data management: Controlling data volume, velocity and variety. *META group research note*, 6(70), 1.

Larranaga, P., Inza, I., & Moujahid, A. (1997). Tema 8. redes neuronales. *Redes Neuronales, U. del P. Vasco*, 12-17.

Lin, C. M., & Gen, M. (2007). An effective decision-based genetic algorithm approach to multiobjective portfolio optimization problem. *Applied Mathematical Sciences*, 1(5), 201-210.

Madahar, A., Ma, Y., & Patel, K. (2017). Application of a Shallow Neural Network to Short-Term Stock Trading. *arXiv preprint arXiv:1703.10458*.

Mahdavi, M., Jin, R., & Yang, T. (2012). Trading regret for efficiency: online convex optimization with long term constraints. *Journal of Machine Learning Research*, 13(Sep), 2503-2528.

Marín Diazaraque, J. M. (2009). Introducción a las redes neuronales aplicadas.

MathWorks. *Deep Learning vs Machine Learning, Ebook*. Obtenido de: https://es.mathworks.com/campaigns/offers/deep-learning-vs-machine-learning-algorithm.html?s_tid=hp_offer_dl_ml_ebok [Visitada en abril 2019]

Matich, D. J. (2001). Redes Neuronales: Conceptos básicos y aplicaciones. *Universidad Tecnológica Nacional, México*.

Moghaddam, A. H., Moghaddam, M. H., & Esfandyari, M. (2016). Stock market index prediction using artificial neural network. *Journal of Economics, Finance and Administrative Science*, 21(41), 89-93.

Navon, A., & Keller, Y. (2017). Financial Time Series Prediction Using Deep Learning. *arXiv preprint arXiv:1711.04174*.

Nordberg, M., & Karlsson, S. (2015). Stock market index prediction using artificial neural networks trained on foreign markets: And how they compare to a domestic artificial neural network.

Park, B., & Van Roy, B. (2015). Adaptive execution: Exploration and learning of price impact. *Operations Research*, 63(5), 1058-1076.

Patgiri, R., & Ahmed, A. (2016, December). Big data: The v's of the game changer paradigm. In *2016 IEEE 18th International Conference on High Performance Computing and Communications; IEEE 14th International Conference on Smart City; IEEE 2nd International Conference on Data Science and Systems (HPCC/SmartCity/DSS)* (pp. 17-24). IEEE.

Ramos, E. G., & Martínez, F. V. (2013). A Review of Artificial Neural Networks: How Well Do They Perform in Forecasting Time Series? *Analítika: revista de análisis estadístico*, (6), 7-18.

Ritter, G. (2017). Machine learning for trading.

Rout, A. K., Dash, P. K., Dash, R., & Bisoi, R. (2017). Forecasting financial time series using a low complexity recurrent neural network and evolutionary learning approach. *Journal of King Saud University-Computer and Information Sciences*, 29(4), 536-552.

SAS. *Aprendizaje automático, Qué es y por qué es importante*. Obtenido de SAS The power to know: https://www.sas.com/es_es/insights/analytics/machine-learning.html [Visitada en mayo de 2019]

Sezer, O. B., Ozbayoglu, M., & Dogdu, E. (2017). A Deep neural-network based stock trading system based on evolutionary optimized technical analysis parameters. *Procedia computer science*, 114, 473-480.

Shanmugamani, R. (2018). *Deep Learning for Computer Vision: Expert techniques to train advanced neural networks using TensorFlow and Keras*. Packt Publishing Ltd.

Sharang, A., & Rao, C. (2015). Using machine learning for medium frequency derivative portfolio trading. *arXiv preprint arXiv:1512.06228*.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, 15(1), 1929-1958.

Stern, D., Samulowitz, H., Herbrich, R., Graepel, T., Pulina, L., & Tacchella, A. (2010, July). Collaborative expert portfolio management. In *Twenty-Fourth AAAI Conference on Artificial Intelligence*.

Su, X. (2012). Introduction to Big Data. *Institutt for informatikk og e-l ring ved NTNU*
Learning material is developed for course IINI3012 Big Data

The Economist. *The Economist Data Tool, Country Risk*. Obtenido de The Economist:
http://graphics.eiu.com/data_services/contentguide/eeiucoun.htm
[Visitada en mayo 2019]