



Facultad de Ciencias Económicas y Empresariales

BIG DATA EN VALORACIÓN. CONCEPTUALIZACIÓN, ANÁLISIS Y APLICACIÓN DEL BIG DATA A DISTINTOS ASPECTOS DE LA VALORACIÓN DE EMPRESAS.

Autor: Guillermo Sáez Jiménez

Director: Leandro Sergio Escobar Torres

Resumen

El Big Data se ha convertido en un elemento clave de nuestra vida cotidiana, tanto en el ámbito económico como social. El conocimiento en detalle de las herramientas capaces de trabajar con este tipo de datos ofrece multitud de aplicaciones, y en este trabajo se explorará en profundidad una de ellas: su uso como herramienta de análisis para la valoración de empresas. Para ello, se realizará un estudio de caso en el que, utilizando distintas herramientas de Big Data, se demostrará la correlación entre el número de veces que aparecen en las redes sociales determinados términos y la evolución de la cotización de las empresas del IBEX 35, el principal índice bursátil español.

Palabras clave: Big Data, valoración, datos, herramientas, análisis, predicción.

Abstract

Big Data has become a key element of our daily life, both in economic and social terms. Knowing in depth how to use the tools that are capable of working with this kind of data offers many applications, and in this article one of them will be thoroughly explored: its use as an analysis tool for company valuation. To this effect, a case study will be conducted, in which by using different Big Data tools, a correlation will be proved between the number of times that certain terms are published in the social media and the evolution of stock prices of the listed companies in the IBEX 35, the main Spanish stock exchange index.

Key words: Big data, valuation, data, tools, analysis, prediction.

Índice

1. Introducción	4
a. Objetivos.....	4
b. Metodología.....	4
c. Estado de la cuestión	5
d. Estructura del trabajo.....	6
2. Marco conceptual	7
a. Big Data.....	7
b. Valoración de empresas	12
Métodos basados en el descuento de flujos de caja	12
Métodos basados en el balance.....	14
Métodos basados en la cuenta de resultados	15
Métodos basados en el fondo de comercio	16
Métodos basados en otros criterios	17
c. Big Data en la valoración de empresas.....	18
3. Datos	22
4. Análisis	26
a. Desarrollo del análisis.....	26
b. Resultados	31
c. Predicciones.....	32
5. Conclusiones	36
6. Bibliografía	39
7. Anexos	45
1. Índice de términos	45
2. Tabla de periódicos más seguidos en Twitter España	46
3. Tabla de tuits por día y hora	47
4. Presencia de cada variable en cada modelo.....	48

1. Introducción

a. Objetivos

El objetivo principal de este trabajo es demostrar la utilidad que tiene el Big Data y sus distintas herramientas en el área de la valoración de empresas. Para ello, se estudiará el caso concreto del uso de dos programas de Big Data como herramientas para analizar la relación entre el contenido en redes sociales y la evolución del principal índice bursátil español. A través de este análisis se determinará si un uso adecuado de estas herramientas puede ser de utilidad para comprender la evolución de los precios de las acciones de una determinada compañía, así como los modelos predictivos que se pueden construir en el caso de desear pronosticar de manera aproximada la evolución futura de estos valores.

Como objetivos secundarios, este trabajo pretende describir en qué consisten el Big Data y la valoración de empresas. En relación con el primer concepto, se exploran elementos como sus dimensiones, las clasificaciones más habituales, sus limitaciones y su valor en el mercado de datos global. Respecto a la valoración de empresas, se describen, sin entrar en demasiada profundidad, los principales métodos de valoración empleados hoy en día para estimar el valor económico de las compañías, así como los posibles usos de estas estimaciones. Además, se realiza un análisis de la combinación de ambos conceptos, es decir, de cómo el Big Data es utilizado como herramienta de valoración de empresas, y qué ventajas y limitaciones presenta esta relación.

Por último, un tercer objetivo de este trabajo es demostrar la importancia que tienen hoy en día las redes sociales en los mercados financieros, y como estas, a pesar de generar un flujo de datos continuo y de enormes dimensiones, suponen una fuente de información de valiosa utilidad para conocer el contexto económico global y realizar predicciones futuras con un elevado grado de precisión y solidez.

b. Metodología

La metodología de este trabajo ha consistido en sustentar una hipótesis, que es la importancia del Big Data como herramienta para valorar una o más compañías, a través de un estudio de caso. Para ello, se comenzó elaborando un marco conceptual, construido a partir de diversas fuentes, de carácter tanto primario (datos oficiales, webs corporativas, periódicos, etc.) como secundario (más de veinte artículos académicos).

Una vez desarrollado el marco teórico, se procedió al estudio de caso como tal, que consistía en estudiar la correlación entre el contenido de los tuits publicados por los principales periódicos en España durante un período determinado de tiempo, y la evolución de la cotización de las empresas que componen el IBEX 35 durante esas mismas semanas. Para obtener los datos relativos a los tuits, se utilizaron dos herramientas: *Vicinitas* para la descarga de los 3200 tuits más recientes de cada uno de los once periódicos, y Excel para el recuento del número de menciones a determinados términos en esos tuits. En cuanto a la obtención de los datos de cotización de las empresas, se descargaron de la web financiera *Investing.com* y se procesaron asimismo en Excel.

Por último, para realizar el análisis estadístico se utilizó el programa informático *Gretl*, especializado en análisis econométricos. En concreto, se realizó un modelo de regresión lineal múltiple para cada una de las 35 compañías del IBEX y otro para el índice de manera global, además de análisis predictivos para aplicar estos modelos a posibles futuros escenarios. Todos los resultados fueron plasmados en la segunda mitad del trabajo, dedicada al estudio de caso, y finalmente se expusieron las conclusiones del análisis y la constatación de la hipótesis de partida.

c. Estado de la cuestión

Hoy en día, el Big Data ha alcanzado tales dimensiones en prácticamente todos los sectores del ámbito económico, político y social que resulta imposible comprender el mundo actual sin tener en cuenta este concepto. Por Big Data se conoce a todo conjunto de datos caracterizado por su gran volumen, variedad, velocidad y veracidad. Estos datos pueden presentarse de forma estructurada o desestructurada, con origen interno o externo, de carácter personal o anónimo, etc. No obstante, a pesar de su diversidad y volumen, disponiendo de las herramientas adecuadas pueden resultar de una gran utilidad, al ser capaces de aportar información de calidad en un periodo de tiempo y con una facilidad nunca antes conocidas.

Uno de los aspectos en los que destaca la utilidad del Big Data y de las herramientas que permiten extraer información valiosa de él es el de la valoración de empresas, que sirve a diversos propósitos y que en los últimos años ha incorporado técnicas de Big Data para poder llevar a cabo estimaciones más precisas y por tanto útiles, así como modelos de predicción cada vez más fiables y manejables. Numerosos estudios e informes han

explorado esta aplicación del Big Data, pero resulta evidente que, dada la naturaleza dinámica y cambiante de este tipo de herramientas, aún existe margen para ampliar el conocimiento sobre las posibilidades que ofrece y que puede ofrecer en el futuro cercano.

d. Estructura del trabajo

Este trabajo se compone de dos mitades. La primera es un marco conceptual que desarrolla de manera teórica los conceptos del Big Data, la valoración de empresas y la combinación de ambos. La segunda mitad comienza con un apartado de datos, en el que se describen las técnicas empleadas para obtener los datos utilizados en el estudio de caso, así como las características más destacadas de estos. Posteriormente, en la sección de análisis, se desarrolla en detalle el estudio de caso, centrado en la correlación entre el contenido en redes sociales y la evolución de los índices bursátiles. Finalmente, se llegan a unas conclusiones finales que responden a las preguntas planteadas en los objetivos y sintetizan los resultados del estudio.

2. Marco conceptual

a. Big Data

Hoy en día, resulta imposible comprender el mundo y la sociedad en los que vivimos sin tener en cuenta un elemento clave de ellos: los **datos**. Los individuos y las empresas estamos constantemente generando datos, de forma consciente o inconsciente, que son procesados y analizados para fines económicos, políticos, de seguridad... (PWC, 2019). Estos datos por si solos presentan una utilidad limitada; deben ser agrupados, procesados y analizados para cumplir con su objetivo principal, que es el de **aportar información de calidad** (Martins et. al, 2020).

En este contexto, cabe diferenciar una categoría de datos que ha adquirido una enorme relevancia en las últimas décadas: el **Big Data**. Con este nombre se conoce a las grandes agrupaciones de datos que integran piezas de información diversas y complejas y que requieren de herramientas especiales para poder ser procesados (Casagrande, 2021). Al hablar de Big Data, existen cuatro **dimensiones principales** que permiten ubicar a un conjunto de datos en esta categoría¹ (Wiener, Saunders y Marabelli, 2020):

1. **Volumen**, entendido como la cantidad de datos capturados y procesados.
2. **Variiedad**, entendida como el rango y la diversidad de categorías de las fuentes de datos.
3. **Velocidad**, que hace referencia a la rapidez o frecuencia con que los datos se registran, actualizan y analizan.
4. **Veracidad**, que refleja la fiabilidad de los datos procesados.

Para que un conjunto de datos, o *data set*, sea considerado Big Data, debe cumplir estos cuatro requisitos simultáneamente. Poca utilidad puede tener un conjunto de datos variado, extenso y actualizado, si la procedencia de esos datos no es fiable o la autenticidad de su contenido es dudosa; del mismo modo, si el mismo *data set* resulta fiable y veraz, pero no es actualizado a un ritmo veloz y frecuente, no puede ser considerado Big Data al incurrir en el riesgo de aportar datos obsoletos o potencialmente sustituibles. No siempre resulta sencillo que un único *data set* cumpla con estos cuatro

¹ El número de dimensiones del Big Data varía en función del autor. La clasificación más amplia, conocida como las 10 Vs, incluye otros elementos como la viralidad (mide la rapidez con la que se propagan los datos por la red), la viscosidad (hace referencia a la facilidad de navegar a través de los datos) o el valor (monetario o de otros tipos). (Moro Visconti, et. al, 2018).

requisitos, y de ahí procede el **valor** del Big Data en el mundo actual (Wiener, Saunders y Marabelli, 2020).

Por otra parte, existen distintas **clasificaciones** del Big Data, atendiendo a factores como el formato de los datos o la fuente de estos (Casagrande, 2021). Con relación a la primera clasificación, se suele distinguir entre datos **estructurados** y **no estructurados**, en función de si existe o no la posibilidad de organizarlos en una base de datos de manera estructurada, por ejemplo, en filas y columnas (Syed et al., 2013). A pesar de que los datos estructurados comportan una mayor facilidad de procesamiento y por ello su análisis suele ser más revelador, más del 80% del Big Data hoy en día se compone de datos desestructurados, y el proceso de convertirlos en datos estructurados puede suponer unos costes muy elevados (Casagrande, 2021).

En cuanto a la clasificación del Big Data según la procedencia de los datos, se suele distinguir entre Big Data **interno** y **externo**. Mientras que el Big Data interno corresponde a datos que maneja la propia entidad interesada en analizar los datos, el externo debe ser obtenido de una fuente ajena a la entidad. Ambas categorías presentan ventajas e inconvenientes. El Big Data interno, que incluye ejemplos como los datos de ventas de una compañía, sus facturas o sus correos electrónicos, suele comportar un menor coste económico y es más rápido y sencillo de obtener; sin embargo, la información que aporta su procesamiento y análisis suele ser menos valiosa que aquella obtenida con el Big Data externo, procedente de fuentes ajenas a la empresa o entidad como las redes sociales, los organismos públicos o las investigaciones de mercado desarrolladas por agencias independientes (Casagrande, 2021).

Además del formato y la procedencia de los datos, existen otras categorías de Big Data que revisten un gran interés, y que pueden presentarse tanto en un formato estructurado como desestructurado, además de obtenerse de forma interna o externa. Ejemplos de estas subcategorías son los datos transaccionales (pagos, cobros, ventas...), el *metadata* (hace referencia a otros datos, facilitando su agrupación y/o análisis) o los datos derivados (se obtienen de combinar, agregar y procesar otros datos) (PWC, 2019). Asimismo, en los últimos años ha cobrado una enorme relevancia el concepto de **datos personales**, que hace referencia a los datos de un individuo en particular, y que pueden ser obtenidos de forma voluntaria si el individuo los comparte conscientemente, estudiada a partir de la observación de comportamientos o inferida de datos preexistentes (Casagrande, 2021). Según un estudio de la Comisión Europea, en 2020 el valor de este

tipo de datos equivale a un billón de euros, es decir, es equivalente al 8% del PIB de la Unión Europea (PWC, 2019).

Por otro lado, los datos, como cualquier otra categoría de activo, representan un valor susceptible de ser medido y comparado. Algunos de los **drivers de valor** más importantes en el área del Big Data son la exclusividad de los datos, su precisión (que no debe ser confundida con su veracidad, ya que un conjunto de datos puede ser fiable pero aportar una información no del todo precisa), sus limitaciones de uso (no todos los datos están disponibles en cualquier momento y de cualquier forma) o los riesgos y responsabilidad que conllevan su obtención y su análisis (PWC, 2019). Cuantificar este valor no es una tarea sencilla, y aunque hoy en día no existe un consenso generalizado sobre la técnica más adecuada para hacerlo, existen tres **métodos** principales.

1. **Enfoque de renta.** Consiste en calcular el incremento en los flujos de caja generados por el uso de los datos. Este incremento puede provenir de un aumento en los ingresos o de una reducción en los costes, y es el resultado del uso que la entidad haya realizado del conjunto de datos.
2. **Enfoque de mercado.** Valora los datos en función del valor que les da el mercado, es decir, el precio al que son comprados y vendidos. Actualmente existen pocos mercados especializados en el comercio de datos, aunque algunos ejemplos son el BDEX, donde los usuarios pueden vender a empresas y agencias datos personales como sus hábitos de compra o sus historiales de búsqueda; o Flickr, donde se compran y venden imágenes digitales.
3. **Enfoque del coste.** Asigna valor a los datos en función del coste que supone para la entidad acceder a ellos y procesarlos. Según este enfoque, los datos aportan valor cuando estos costes son inferiores a los ingresos que de estos datos se obtienen, es decir, cuando aportan rentabilidad (PWC, 2019).

Además, en el ámbito empresarial, las compañías pueden maximizar el valor de sus *data sets* de distintas maneras: optimizando sus procesos internos, mejorando sus productos y servicios, monetizando los datos a través de su venta a terceros... Cuando uno o varios de estos usos del Big Data suponen la actividad económica principal de la compañía que los utiliza, se habla de “**Modelos de Negocio Big Data**”, o BDBM, por sus siglas en inglés. Estos modelos de negocio, a su vez, pueden clasificarse según el uso que hacen del Big Data en usuarios de datos, proveedores de datos o facilitadores de datos,

que proveen a los dos primeros de infraestructura de datos y/o servicios relacionados con el Big Data (Wiener, Saunders y Marabelli, 2020).

Este último modelo de negocio tiene su razón de ser en el elevado coste que conllevan a menudo las infraestructuras necesarias para obtener y procesar Big Data, y que provocan que las empresas opten por externalizar estos procesos. Estructuras como los **almacenes de datos** (más conocidos por el nombre de *data warehouses*), en las que los datos son preparados, diseñados y almacenados, o herramientas como la minería de datos y el procesamiento analítico en línea (OLAP, por sus siglas en inglés) requieren de una inversión que a menudo está justificada únicamente en aquellas ocasiones en las que el uso del Big Data está contemplado como una herramienta a largo plazo, y no cuando se requiera de una manera puntual (Martins et. al, 2020). Por ello, numerosas empresas tecnológicas como IBM y SAS han comenzado a comercializar programas de análisis de Big Data a sus clientes (Fanning y Drog, 2014).

Otro inconveniente de considerable interés a la hora de obtener y analizar Big Data que empuja a las empresas a externalizar estos procesos es el respeto a la normativa y la legislación que afectan a esta actividad en cada mercado donde opere la compañía. Esta regulación es particularmente restrictiva en el contexto comunitario europeo, especialmente desde la aprobación en 2016 del **Reglamento General de Protección de Datos**, comúnmente conocido por sus siglas en inglés GDPR. Esta normativa ha obligado a un gran número de empresas a dedicar considerables esfuerzos económicos y organizativos a cumplir con los requisitos que impone al tratamiento de datos personales, que se agrupan principalmente en cuatro principios: transparencia, integridad, confidencialidad y autenticidad (Martins et. al, 2020). Con el objetivo de minimizar los riesgos legales que supone la obtención y el tratamiento de datos, cada vez más empresas optan por subcontratar estas tareas a otras empresas (Pérez, 2021).

Existen, además, algunos aspectos del Big Data que pueden suponer una verdadera **amenaza** para las empresas hoy en día. El ejemplo más evidente es el de la **ciberseguridad**, puesto que los ataques informáticos aumentan su eficacia y su nocividad en función del volumen y el valor de los datos que utilizan (Hurtado de Mendoza, 2020). Otro efecto perjudicial que puede tener el Big Data para una compañía es su **huella reputacional** en redes sociales, ya que estas pueden convertirse en un arma de doble filo al posibilitar que un mensaje o discurso se propaguen a una gran velocidad y lleguen a un número de gente mayor que nunca. Esta información puede ser beneficiosa o dañina para

la reputación de la empresa pero, en todo caso, una vez que se libera en las redes sociales escapa el control de la compañía, y por ello debe ser tratada con precaución (Selby, 2015).

Sin embargo, a pesar de estos inconvenientes, lo cierto es que un uso adecuado del Big Data puede aumentar exponencialmente el valor de una compañía, a través de distintos fenómenos como los denominados **efectos de red** (*network effect*) y de **plataforma** (*platform effect*). El primero hace referencia al impacto que tienen los usuarios de un servicio en el valor que este servicio tiene para el resto de usuarios; es decir, cada usuario es al mismo tiempo consumidor del servicio y también proveedor, ya que el valor añadido de este servicio es precisamente la interacción y la conexión entre distintos usuarios. Ejemplos de este efecto de red son Facebook, Twitter, Instagram y LinkedIn. Por otro lado, el efecto de plataforma es similar, aunque se centra más en el valor de la información que se comparte entre los usuarios a través de la plataforma, y no tanto en la conexión entre ellos, como es el caso de la plataforma de vídeos YouTube (Casagrande, 2021).

Así pues, en vista de las numerosas utilidades y efectos positivos del Big Data, no es de extrañar que los modelos de negocio basados en este ámbito sean cada vez más numerosos y que su valor en el mercado esté en auge. Según un análisis de Statista, el **mercado global** de Big Data ha pasado de los 7.600 millones de dólares en 2011 a superar los 50.000 millones en 2020, y se espera que esta cifra alcance los 100.000 millones en 2027 (Statista, 2018). Este enorme crecimiento ha afectado principalmente a los **sectores** tecnológicos, de telecomunicación y financieros, y ha provocado que los datos y la información hayan pasado a ser considerados por muchos expertos como el “**nuevo petróleo**” de la economía global (Casagrande, 2021).

Para finalizar, no obstante, cabe destacar que estas cifras no solo reflejan la proliferación de modelos de negocio basados en el Big Data, sino también el valor que este adquiere en transacciones o actividades puntuales, como pueden ser los estudios de mercado, el diseño del CRM de una empresa o la **valoración de una compañía** con anterioridad a un proceso de fusión o adquisición. Este último aspecto será desarrollado en profundidad en el apartado c, y para ello se comentará previamente en qué consisten estos procesos y cuáles son las principales técnicas de valoración utilizadas hoy en día.

b. Valoración de empresas

De entre las numerosas aplicaciones que tiene el uso de Big Data en el ámbito empresarial, una de las más relevantes debido a su gran utilidad es la valoración de empresas. Existen distintos **propósitos** que pueden motivar la valoración de una compañía: la estimación del precio razonable en una operación de compraventa, el análisis de inversiones en bolsa (se suele comparar el valor de la compañía con su cotización en el mercado), salidas a bolsa, procesos de arbitraje, análisis de *drivers* de valor de una empresa o sector, decisiones estratégicas como el desarrollo de nuevos productos y/o mercados... (Fernández, 2008).

Al ser una actividad que responde a tan diversos propósitos y que se puede aplicar a cualquier tipo de compañía, existen distintos **métodos** de valoración. La parte interesada en valorar una empresa se decantará por uno u otro en función de sus necesidades y del contexto en el que se desarrolle la valoración, ajustándose así al enfoque que más desee adoptar de entre las tres categorías enunciadas en la página 7: enfoque de renta, de mercado o de costes (IVSC, 2022). Así pues, los métodos de valoración más utilizados actualmente son los siguientes²:

1. Métodos basados en el descuento de flujos (*cash flows*).
2. Métodos basados en el balance.
3. Métodos basados en la cuenta de resultados.
4. Métodos basados en el fondo de comercio (*goodwill*).
5. Métodos basados en otros criterios.

Métodos basados en el descuento de flujos de caja

En primer lugar, los métodos basados en el descuento de *cash flows* están orientados a determinar el valor de la compañía estimando los **flujos de fondos** que esta generará **en el futuro**, y **descontando** estos flujos a una determinada **tasa** en función del riesgo que estos acarreen. El objetivo de este descuento es determinar el valor actual de la compañía, ya que los flujos tendrán en un lugar en un momento futuro. Este método es actualmente el más utilizado en el ámbito de la valoración, ya que tiene en cuenta las

² La siguiente clasificación de métodos de valoración, así como su descripción pormenorizada, están inspirados en el libro “Métodos de valoración de empresas”, de Pablo Fernández (2008).

expectativas que se tienen de la compañía a largo plazo (asumiendo siempre su continuidad).

La ecuación básica de este método es la siguiente:

$$V = \frac{CF_1}{1 + K} + \frac{CF_2}{(1 + K)^2} + \frac{CF_3}{(1 + K)^3} + \dots + \frac{CF_n + VR_n}{(1 + K)^n}$$

donde CF_i equivale al *cash flow* generado en el periodo i , K equivale a la tasa de descuento que refleje adecuadamente el riesgo de los *cash flows* y VR_n equivale al valor residual de la compañía en ese periodo (es decir, el valor que resultaría de liquidar la compañía en ese momento).

Otro procedimiento más **sencillo** es considerar una duración indefinida de los flujos después del periodo n y asumir un crecimiento constante (g) de estos flujos. El mayor inconveniente de este método es la necesidad de determinar este crecimiento constante de manera precisa. Si la parte valoradora considera que este método es el más adecuado, la ecuación que debe utilizar es la siguiente:

$$VR_n = \frac{CF_n \times (1 + g)}{k - g}$$

Asimismo, en este método de valoración es necesario especificar qué se entiende por **flujo de caja**, ya que existen **distintas interpretaciones** de este concepto en función de si se orientan en la deuda, en el capital o en otros posibles flujos. No obstante, el **flujo de fondos libre** (*free cash flow*) es el más empleado, ya que combina el valor de la deuda con el de las acciones, es decir, refleja el valor total de la compañía. Este tipo de flujos es el generado por la actividad operativa de la compañía después de impuestos, y no tiene en cuenta la deuda financiera.

Para calcularlo, se debe seguir el siguiente procedimiento:

$$FCF = \text{BAIT} - \text{Tasa impositiva sobre BAIT} + \text{Amortización} +/- \text{CAPEX} +/- \text{Incremento de NOF,}$$

donde BAIT es el beneficio antes de intereses e impuestos, CAPEX es la inversión en activos fijos y NOF son las necesidades operativas de fondos.

Por último, para calcular el valor de una compañía utilizando el método del *free cash flow*, la tasa de descuento que se debe utilizar es el **coste medio ponderado del capital**, también denominado **WACC** (*weighted average cost of capital*), que se calcula de la siguiente manera:

$$\text{WACC} = \frac{(E \times K_e) + D \times K_d \times (1 - T)}{E + D}$$

donde D equivale al valor de la deuda, E equivale al valor de mercado de las acciones, E+D es el valor total de la compañía, K_e es la rentabilidad exigida a las acciones, K_d es el coste de la deuda antes de impuestos y T es la tasa impositiva. Como se observa en la fórmula, el cálculo del WACC tiene en cuenta la estructura financiera de la empresa, ya que se ponderan el peso que tienen el coste de la deuda y el de las acciones, respectivamente.

Así pues, el método de valoración de empresas basado en el descuento de flujo de fondos libre consiste en determinar el valor de este flujo, su tasa de descuento apropiada (WACC), el valor residual y finalmente actualizar los flujos para determinar el valor de negocio actual. Este valor de la empresa hace referencia a la suma del valor de los recursos propios sumado al valor de la Deuda Financiera Neta. No obstante, existen otros cuatro grandes grupos de métodos para calcular este valor, que serán comentados brevemente a continuación.

Métodos basados en el balance

En segundo lugar, los métodos basados en el balance vinculan el valor de una empresa con el valor de su **patrimonio**, es decir, con el **activo** reflejado en su balance de situación. Existen cuatro versiones de este modelo:

1. **Método del valor contable.** El valor de la compañía equivale al valor contable (o en libros) de sus acciones, es decir, al valor de los recursos propios de su balance de situación, que se puede calcular fácilmente sustrayendo el pasivo exigible al activo de la compañía.

2. **Método del valor contable ajustado.** Es similar al método anterior, aunque ajustando el valor del activo y del pasivo de la compañía a sus respectivos valores de mercado. De esta manera se minimiza la problemática de aplicar únicamente criterios contables.
3. **Método del valor de liquidación.** Considera que el valor de una compañía equivale al valor resultante de su liquidación, es decir, el valor de la compañía después de vender sus activos y saldar sus deudas. Se calcula restando los gastos de liquidación (como las indemnizaciones por despido) al patrimonio neto ajustado al valor de mercado. Emplear este método supone asignar a la compañía su valor mínimo, por lo que ofrecerá un resultado inferior al de cualquier otro método.
4. **Método del valor sustancial.** Determina el valor de una compañía equiparándolo con la inversión que supondría constituir otra compañía de características similares. No tiene en cuenta las partidas del activo que no están relacionadas con la actividad principal de la empresa, como los inmuebles sin explotar o las participaciones en otras compañías.

El principal inconveniente de esta categoría de métodos es que reflejan una **realidad estática** de la compañía, sin tener en cuenta aspectos dinámicos de relevancia como el valor temporal del dinero o la posible evolución futura de la compañía, del sector en el que opera o de la economía en general. Por tanto, su utilidad se limita a valoraciones puntuales de una compañía y no tanto al valor potencial que esta pueda tener en un momento futuro.

Métodos basados en la cuenta de resultados

Los métodos basados en la cuenta de resultados, por su parte, utilizan **múltiplos** para determinar el valor de una empresa a través del volumen de determinadas variables que aparecen en este estado financiero, como los beneficios o las ventas. Estos múltiplos han de ser consistentes y fiables, y solo serán de gran utilidad cuando sirvan para comparar a compañías de características similares. Los principales métodos de esta categoría son los siguientes:

1. **Métodos basados en el PER.** Determinan el valor de las acciones de una compañía multiplicando el beneficio neto anual por una ratio precio-beneficio

denominada PER (*price to earnings ratio*). Esta ratio relaciona la cotización en bolsa de una acción con el beneficio por acción (BPA) de la compañía. El PER es la ratio más empleada en el área de la valoración, especialmente para aquellas compañías que cotizan en la bolsa de valores, y por ello estos métodos son bastante habituales.

2. **Métodos basados en el dividendo.** Determinan el valor de las acciones de una compañía en función del valor de los dividendos que se espera que estas generen. De esta forma, el valor de una acción equivaldría a dividir el dividendo por acción repartido entre la rentabilidad exigida a dicha acción, también conocida como el coste de los recursos propios.
3. **Métodos basados en múltiplos de las ventas.** Determinan el valor de una compañía multiplicando su cifra de ventas por un valor concreto. Esa cifra debe definirse atendiendo a criterios específicos, que suelen tratarse del contexto del sector de la compañía o de la coyuntura económica general.
4. **Métodos basados en otros múltiplos.** Sirven para comparar empresas similares a través de otros valores de su cuenta de resultados. Los múltiplos más utilizados son los resultantes de dividir el valor de la empresa entre su BAIT, el valor de la empresa entre su EBITDA, el valor de la empresa entre su flujo de caja operativo y el valor de cotización de las acciones entre su valor contable.

Métodos basados en el fondo de comercio

En cuarto lugar se encuentran los métodos basados en el fondo de comercio o *goodwill*. Este término hace referencia a la capacidad que tiene una compañía de generar beneficios en el futuro a partir de **activos intangibles** como las marcas con las que comercializa sus productos, la calidad de su cartera de clientes, sus patentes y propiedad intelectual o su posicionamiento en un mercado concreto (Pariente y Reig, 2018). Estos elementos no suelen estar representados en el balance de situación de las compañías o en ningún otro estado financiero, y sin embargo pueden llegar a ser de gran relevancia a la hora de valorar una empresa en su conjunto. Por ello, los métodos de valoración basados en el fondo de comercio tratan de determinar el valor total de una compañía combinando su valor patrimonial con su capacidad intangible de generar beneficios en el futuro.

Métodos basados en otros criterios

Por último, debido a la gran variedad de contextos en los que una empresa puede ser valorada y a las diversas características que esta compañía puede presentar, existen otros criterios de valoración más específicos que en ocasiones resultan más adecuados que los descritos anteriormente. Uno de los más destacados es el del *break-up value*, que consiste en determinar el valor de una compañía evaluando por separado cada una de sus unidades de negocio o sus divisiones y unificando posteriormente estos valores en una suma total. Este método es particularmente útil para grupos industriales muy diversificados o para empresas multinacionales con características muy diferentes en cada territorio. (Fernández, 2008).

Otros métodos se centran en la creación de valor de la compañía, como el método del **Valor Económico Agregado (EVA)**, cuyo objetivo es determinar el beneficio económico real de una compañía teniendo en cuenta aspectos como la confianza de los accionistas en beneficios futuros o la productividad de todos los factores de producción de la compañía, incluidos los intangibles (Bonilla, 2010); así como el método del **CFROI** (*Cash flow return on investment*), que hace referencia a los flujos de caja que una compañía genera en un periodo determinado como porcentaje del efectivo invertido en los activos de esa compañía (Erasmus y Lambrechts, 2006). Finalmente, algunos métodos de valoración se centran en la evaluación **de opciones reales de inversión futuras**, poniendo el foco principal de la valoración en el factor riesgo (Fernández, 2008).

En conclusión, existen múltiples modelos y técnicas para valorar una compañía, sea cual sea el objetivo de esta valoración, y cada uno de ellos tiene en cuenta datos de distinta naturaleza. Por ello, es de gran utilidad disponer de herramientas que permitan procesar estos datos y desarrollar modelos que combinen diferentes técnicas de valoración y presenten esta información de forma que se facilite la toma de decisiones y se minimice el riesgo de estas. En el siguiente apartado se describirá cómo el **Big Data** es utilizado cada vez más a menudo como pilar fundamental **en la valoración de empresas**, y qué posibles aplicaciones en este campo cabe esperar durante los próximos años.

c. Big Data en la valoración de empresas

Una vez descritos y desarrollados los marcos teóricos del Big Data y de la valoración de empresas, resulta apropiado analizar el resultado de la combinación de ambas materias. Como se ha explicado en el primer apartado, actualmente el Big Data resulta fundamental en prácticamente cualquier área del mundo empresarial, y esto incluye desde hace años la valoración de empresas. Para algunos autores, este cambio no ha supuesto únicamente una mejora en los resultados de la gestión empresarial, sino que ha alterado asimismo su orientación general. Tradicionalmente, la dirección estratégica se basaba en un enfoque de arriba abajo (*top-down*), en el que la compañía fijaba unos objetivos y trabajaba para conseguirlos. Sin embargo, desde la aparición del Big Data, cada vez más empresas optan por **estrategias de abajo arriba** (*bottom-up*), basadas en información procedente de la **recolección y el análisis de datos** (Moro Visconti et. al, 2018).

Estos datos son con frecuencia la **base de la valoración** de determinados elementos de una compañía, como su fondo de comercio, su capacidad para capturar cuota de mercado o su promedio de ingresos por usuario (ARPU, por sus siglas en inglés), particularmente útil a la hora de valorar empresas tecnológicas (Selby, 2015). Además, debido a la creciente facilidad y al abaratamiento de los procesos de captación y almacenamiento de datos, cada vez es más sencillo utilizar información de gran relevancia sobre la compañía que se pretende valorar, desde sus sistemas de planificación de recursos empresariales (ERPs) y de gestión de relaciones con el cliente (CRM) hasta sus datos contables detallados (Fanning y Drogt, 2014).

Evidentemente, la facilidad de acceso a esta información dependerá del grado de privacidad que presente y de la relación entre la empresa valorada y la parte valoradora, que puede caracterizarse por una mayor o menor transparencia. No obstante, esta **limitación** no es una novedad propia de la era tecnológica actual, sino que ha existido desde siempre en el área de la valoración de empresas, y, en todo caso, resulta evidente que el Big Data ha contribuido a minimizarla al ir más allá del almacenamiento de datos y permitir diseñar modelos de procesamiento y análisis de la información con los que se hace posible construir una valoración más precisa de la compañía en cuestión (Martins et. al, 2020).

Por otra parte, el Big Data ha permitido **reducir** o incluso eliminar algunos de los **errores** más frecuentes en la valoración de empresas, como las previsiones de flujos de

caja futuros inconsistentes con el contexto económico general o el análisis competitivo del sector (Fernández, 2008). Al disponer de un conjunto de datos de enorme volumen, variedad y veracidad, la parte valoradora es capaz de construir modelos de análisis predictivo mucho más fiables que los desarrollados hace unas pocas décadas, y estimar así los flujos de caja futuros de la compañía valorada con una precisión mucho mayor. Asimismo, a la hora de **identificar** posibles *targets* para una fusión o una adquisición, el uso del Big Data ofrece una herramienta de información mucho más sólida y objetiva que los datos limitados e impulsos subjetivos que solían determinar estos procesos hasta hace no demasiado tiempo (Fanning y Drogt, 2014).

Atendiendo al primer gran beneficio del uso de Big Data, existe un considerable número de elementos de la valoración cuya evaluación resulta bastante más precisa si se utilizan estas técnicas y herramientas. En primer lugar, a la hora de determinar el **EBITDA** de una compañía, el Big Data permite actualizar sus costes operativos variables a tiempo real, así como agregar y clasificar sus ingresos netos de múltiples maneras. Además, el Big Data permite construir modelos de análisis predictivo para pronosticar con gran precisión futuros valores del **WACC**, el CAPEX y las necesidades operativas financieras, que resultan claves para el cálculo de los flujos de caja libres. Por último, la predicción de la **Deuda Financiera Neta** en periodos futuros se puede ver facilitada en gran medida al vincularla con la evolución del CAPEX y del activo circulante a través de modelos de Big Data (Moro Visconti et. al, 2018).

Por otro lado, la segunda gran utilidad del Big Data en la valoración de empresas corresponde al área de las **fusiones y adquisiciones**. Son numerosas las motivaciones que una empresa puede tener para fusionarse con otra o absorberla: crear sinergias, alcanzar el liderazgo en el mercado, acelerar el crecimiento de la compañía e incluso aprovechar una infravaloración de una compañía por parte del mercado, entre otras (Zhang, 2020). En todas ellas, no obstante, se hace imprescindible contar con herramientas que permitan valorar de forma adecuada a la otra compañía, especialmente en el contexto de la adquisición, donde el escenario ideal para la empresa compradora es poder acordar un precio de adquisición inferior al valor real de la empresa adquirida (Fernández, 2008).

Según un estudio de Deloitte, las **herramientas** de Big Data más utilizadas por las empresas durante los procesos de fusiones y adquisiciones son el análisis de consumidores y mercados, la identificación y evaluación de sinergias y el análisis de la fuerza laboral y su remuneración (Deloitte, 2013). Además, una técnica cada vez más

habitual de identificar posibles targets es utilizar las **redes sociales**, que ofrecen un gran volumen de datos a tiempo real sobre aspectos de la compañía como la satisfacción del cliente o de los empleados, e incluso permiten a la parte compradora adelantar en parte el proceso habitual de *due dilligence* característico de los procesos de M&A al permitirle averiguar datos sobre las operaciones de la compañía adquirida (Fanning y Droggt, 2014). En ambos casos, la información recabada de las redes sociales puede ser obtenida y procesada de manera mucho más eficiente a través de un uso adecuado del Big Data.

El Big Data, sin embargo, puede ser de utilidad para más partes interesadas, además de para la empresa compradora. A fin de cuentas, un proceso de fusión o adquisición afecta a todos los *stakeholders* de ambas compañías, desde sus empleados, directivos y accionistas hasta sus acreedores y fondos de inversión. Es por ello por lo que predecir futuros procesos de M&As, que dista de ser una actividad nueva en los ámbitos de la inversión y los mercados de valores, se ha convertido en otra de las utilidades del Big Data en relación con la valoración de empresas. Por ejemplo, dado que históricamente las acciones de una compañía adquirida tienden a aumentar, y las de una compañía compradora suelen caer, sería interesante para un inversor conocer de antemano cuándo va a tener lugar este proceso de adquisición para poder anticiparse así al mercado y obtener cierto rédito de la operación (Moriarty et. al, 2019).

Asimismo, es necesario destacar que todas estas ventajas y utilidades del Big Data en la valoración de empresas no son excluyentes respecto a las **técnicas tradicionales** de valoración. Más bien al contrario, puesto que un modelo de valoración óptimo debería ser capaz de combinar métodos clásicos de valoración como el descuento de flujos de caja libre o los múltiplos sobre determinados valores con otros más innovadores basados en el Big Data, como los análisis de series temporales que integren datos transversales o los modelos predictivos de carácter estocástico (Moro Visconti et. al, 2018). Estos últimos modelos, que atribuyen una distribución aleatoria a determinados valores y configuran patrones de predicción para llevar a cabo estimaciones dentro de un intervalo de confianza concreto, requieren de un volumen de datos tan grande que únicamente pueden configurarse utilizando herramientas de Big Data (Cárdenas y Morales, 2016).

Por supuesto, al igual que con el resto de prácticas relacionadas con el Big Data, las fusiones y adquisiciones motivadas o influidas por estas herramientas han de cumplir con la **legislación** de cada mercado y, lo que a menudo resulta más complejo, han de evitar ser percibidas por los consumidores e inversores como poco éticas o transparentes (Selby,

2015). De no ser así, la operación de fusión o adquisición puede perjudicar gravemente a la **reputación** de las empresas implicadas, llegando incluso a provocar que el resultado final de la operación sea negativo. Un ejemplo de esta situación que alcanzó una gran relevancia por sus repercusiones mediáticas y legales fue la absorción de WhatsApp por Facebook en 2014 y la polémica que esta operación generó en relación con el tratamiento que la compañía de Zuckerberg podría hacer de los datos de los usuarios de la app de mensajería (Otto, 2016).

En todo caso, siempre y cuando se lleve a cabo una buena **planificación estratégica** que tenga en cuenta estos posibles inconvenientes, el uso del Big Data facilita enormemente la tarea de valorar empresas y la toma de decisiones relacionadas con ella, especialmente en los procesos de fusiones y adquisiciones. No obstante, no se debe dejar de lado el **componente humano** de cualquier decisión que afecte a la estrategia global de una compañía, ya que las empresas están formadas por personas, y las decisiones fundamentales que las afecten nunca deben recaer únicamente en un algoritmo o un modelo de procesamiento de datos (Fernández, 2008).

En conclusión, tras analizar en detalle el ámbito del Big Data, las técnicas más habituales de valoración de compañías y cómo lo primero puede aplicarse a los métodos y técnicas propias de lo segundo, se observa que hoy en día una buena valoración puede y debe incluir herramientas que le permitan procesar conjuntos de datos de gran volumen, variedad, velocidad y veracidad. El estudio de caso que compone la segunda mitad de este trabajo se centrará en un elemento fundamental en la valoración de una compañía, su precio de cotización, y lo vinculará con el contenido en redes sociales a través de dos herramientas distintas de Big Data. Con ello se pretende demostrar la importancia que tienen este tipo de instrumentos para realizar estimaciones precisas del valor de una compañía, así como para tratar de predecir su evolución en el futuro.

3. Datos

Antes de comenzar a explicar el desarrollo del estudio de caso y los resultados obtenidos en él, es necesario describir el proceso de **obtención** de los datos empleados para el análisis, especialmente dada la temática general de este trabajo. Así pues, para llevar a cabo el siguiente estudio, ha sido necesario obtener datos de distinta naturaleza, que se dividen en dos categorías principales: los datos históricos de cotización del IBEX 35 y de cada una de las 35 empresas que lo componen durante el mes previo al análisis, y el número de tuits generados durante esas semanas que contuvieran determinados términos en su contenido.

En primer lugar, la obtención de los **datos históricos diarios de cotización** del IBEX 35 no presentó grandes dificultades, puesto que estos están disponibles en numerosas páginas webs de análisis financiero. En concreto, para la realización de este trabajo, se descargaron los datos de la página *Investing.com*, que ofrecía un análisis diario de valores diarios como los precios de apertura y cierre, los valores máximos y mínimos de la jornada y la evolución respecto al día anterior. El único inconveniente surgido en el proceso fue la exportación de las cotizaciones diarias a Excel, puesto que aparecían copiados por defecto todos los datos en la misma columna, y hubo que llevar a cabo a una clasificación por columnas de cada una de las 35 empresas, de forma que se pudiera trabajar con cada uno de los datos por separado. Una vez finalizado este proceso, se unificaron en un único documento de Excel todas las cotizaciones, dispuestas en 36 hojas (la primera para el IBEX en conjunto, y el resto para cada una de las 35 compañías que componen el selectivo).

En cuanto al tratamiento de los **tuits**, que representan la **variable independiente** del análisis, el procedimiento resultó bastante más complejo y dilatado, con ciertos inconvenientes imprevistos. En primer lugar, se llevó a cabo un estudio comparativo de las distintas herramientas de **scraping** de redes sociales, es decir, herramientas de Big Data capaces de extraer grandes cantidades de datos de interés de distintas plataformas para facilitar su análisis. En un principio se optó por utilizar el programa informático *Octoparse* ⁸, puesto que, a pesar de requerir numerosas acciones para acceder a tuits a gran escala, se trataba de una herramienta capaz de trabajar con grandes cantidades de datos simultáneamente, por lo que resultaba idónea para el propósito del trabajo: acceder a miles de tuits rápidamente y clasificarlos en función de su contenido.

Sin embargo, a medida que el programa era ejecutado, surgieron inconvenientes de distinta naturaleza. En primer lugar, *Octoparse 8* opera recargando una página web durante un elevado número de veces, permitiendo obtener de manera extremadamente rápida un gran conjunto de datos que, de ser localizados de manera manual, conllevaría una cantidad de tiempo considerable. Sin embargo, puesto que el diseño de plataforma de Twitter está programado para que al cargar de nuevo la página no se acceda a contenido más antiguo, sino que se actualice el contenido novedoso más “interesante” para el usuario, con *Octoparse 8* resultaba imposible acceder a tuits cuya antigüedad fuera superior a los cuatro o cinco días. Esto haría inviable cualquier análisis de correlación sólido, ya que, en el mejor de los casos, se podría comparar el contenido de los tuits con la cotización del IBEX 35 durante una única semana, y además existiría una gran diferencia entre el número de datos relativos al día de la descarga y el resto de las jornadas.

Fue así como se decidió sustituir la idea original de indagar en todo el contenido de Twitter por la de centrarse en los perfiles de los diez **periódicos** con más seguidores de Twitter España. Sin embargo, una vez realizada esta modificación en el planteamiento del análisis, surgió una nueva problemática. Para algunos periódicos, como *El País* o *ABC*, resultaba imposible acceder a tuits con más de dos semanas de antigüedad. Ante esta situación, se procedió a repetir el análisis con otras herramientas de *scraping*, con el objetivo de comprobar si el problema era causa de *Octoparse 8*. Al obtener resultados similares con otros dos programas, se llevó a cabo una nueva investigación en Internet y se constató que Twitter únicamente almacena los últimos 3200 tuits de cada perfil, de forma que para acceder a los anteriores sería necesario crear un *data set* y actualizarlo para su consulta posterior. Esta limitación afectaba principalmente a la obtención de tuits procedentes de perfiles de periódicos generalistas, que publican un volumen considerable de tuits al día (*El País*, por ejemplo, emitía un promedio de 190 tuits diarios).

Además, durante esta comprobación, se constató asimismo que el programa *Octoparse 8* no resultaba ser una herramienta tan útil como se había considerado al principio del proceso, puesto que su uso era demasiado complejo y finalmente el análisis no iba a requerir un volumen de datos tan elevado (unos 35000 tuits, frente a los millones que se esperaban obtener al principio del estudio). Así pues, a partir de ese momento, se empleó la herramienta online *Vicinitas*, especializada en la extracción de datos de Twitter. Con ella se importaron a Excel los 3200 tuits más recientes de los once periódicos seleccionados para el trabajo. Entre estos diarios, se incluían seis de carácter generalista

(*El País, El Mundo, ABC, la Vanguardia, elDiario y El Confidencial*) y otros cinco especializados en finanzas (*Expansión, elEconomista, Cinco Días, BolsaManía y Cotizalia.*).

Los criterios para seleccionar qué periódicos habían de ser incluidos en el análisis se basaron, en el caso de los periódicos generalistas, en el número de seguidores, puesto que los perfiles de Twitter de esos seis diarios eran los que más seguidores presentaban de todo el panorama nacional (ver Anexo 6.2). En el caso de los periódicos financieros, por su parte, el objetivo de incluirlos en el análisis respondía a la premisa de que, debido a su área de especialidad, el contenido de sus tuits revestiría de un gran interés para los inversores del selectivo español. Al igual que en el caso de los periódicos generalistas, se procedió a seleccionar aquellos perfiles que contaran con un número mayor de seguidores en Twitter, dentro de su área informativa.

Así pues, una vez descargados los 3200 tuits más recientes de cada periódico (a fecha de 19 de marzo), se procedió a clasificar estos tuits según su **contenido**. En concreto se seleccionaron diez **términos** relacionados con el conflicto entre Ucrania y Rusia: Ucrania, Putin, Rusia, Guerra, Invasión, Biden, Ejército, OTAN, Kiev y Crisis. Gracias a la función “Contar” de Excel, fue posible cuantificar el número de veces que se repetían diariamente cada uno de esos términos, en cada uno de los once periódicos. Puesto que la inmensa mayoría de los tuits resultantes correspondían a las fechas posteriores al 18 de febrero, y esa fecha coincidía con las cuatro semanas previas a la descarga de los datos de cotización, se decidió limitar el estudio a esos 30 días, que en la práctica serían 21, puesto que los fines de semana la bolsa no opera, y por tanto los tuits publicados en sábados y domingos no serían tenidos en cuenta para el análisis.

Se llegó así a una cifra final de **15.124 tuits**, de los cuales el 97% (14.731) correspondían a las cuatro semanas previas a la descarga de los datos. Además, en numerosos casos un mismo tuit contenía dos o más términos destacados, por lo que el número total de palabras detectadas fue de más de algo más de 21.000. Evidentemente, un análisis más exhaustivo habría tenido en cuenta posibles variaciones de estos términos, tales como gentilicios (“ruso”, “rusa”, “rusos”) o sinónimos y expresiones equivalentes (“ejército” por “tropas”, “Kiev” por “capital”). No obstante, para el propósito de este trabajo, se consideró que un total de **21.000 términos** serían suficientes para demostrar la posible correlación entre su aparición en Twitter y la respuesta de los mercados financieros.

Por otro lado, esta cifra total contenía una ostensible variación entre valores de una jornada u otra. Así, por ejemplo, el 24 de febrero, día en el que comenzaron las agresiones armadas a Ucrania por parte de Rusia, únicamente se registraron 370 tuits conteniendo los términos seleccionados. No obstante, al día siguiente, esta cifra se multiplicó por seis, al alcanzarse los 2.305 tuits, y este dato continuó aumentando hasta llegar a los 3882 registrados el martes 1 de marzo³. Asimismo, era posible apreciar una notable división entre los cinco términos más recurrentes (Ucrania, Putin, Rusia, Guerra e Invasión), y los cinco restantes. Así, mientras que los nombres de ambos países y del dirigente ruso aglutinaban miles de tuits (en el caso de Ucrania, 7160), palabras como Biden, Ejército y OTAN apenas llegaban a los 500 en total.

No obstante, esta dispersión de resultados, especialmente en relación con las fechas, resultaba de gran interés para el análisis, puesto que al presentar un mayor rango de valores, permitiría demostrar con mayor exactitud la posible **correlación** entre las dos variables del estudio (cotización y tuits). Por ello, una vez estructurados y clasificados todos los datos, se seleccionaron los más relevantes (número total de tuits por palabra y fecha), y estos fueron unificados con los procedentes del análisis histórico del IBEX 35 en un único archivo Excel. Este archivo final sería el utilizado posteriormente para exportar los datos a *Gretl*, un programa informático especializado en modelos econométricos con el que se desarrollaría el análisis de regresión lineal.

³ Las cifras correspondientes a cada día y palabra están disponibles en el Anexo 6.3.

4. Análisis

a. Desarrollo del análisis

El análisis práctico de este trabajo ha sido desarrollado, además de con Excel y con las herramientas de extracción de datos de redes sociales, con el programa informático *Gretl*. Este software de libre acceso está orientado al **análisis econométrico**, y por ello ha sido seleccionado para estudiar la relación entre las variables del análisis. En concreto, para comprobar si existe o no relación entre la cotización diaria del IBEX 35 y de sus empresas y la aparición de determinados términos en las redes sociales, se ha optado por llevar a cabo un **análisis de regresión lineal múltiple**. Este tipo de análisis tiene como objetivo calcular la correlación entre una variable dependiente y dos o más independientes, también llamadas regresores, a través de una ecuación lineal.

En este análisis se ha optado por utilizar como variable dependiente la diferencia entre la cotización al cierre de una jornada y la cotización al cierre de la jornada anterior. A esta variable se le denominará en adelante *Diferencial_n*. Además, para hacer el análisis econométrico más preciso, esta variable se obtiene a partir de la diferencia de los **logaritmos** naturales de cada cotización, es decir:

$$\text{Diferencial}_{n1} = \ln(\text{diferencia}_1) - \ln(\text{diferencia}_0)$$

La razón de utilizar logaritmos en este análisis econométrico es minimizar el número de valores atípicos y suprimir el impacto de las unidades de las variables (en este caso, los euros) sobre los cocientes estimados, entre otras ventajas. En todo caso, los datos descargados de *Investing.com* incluyen otras **variables** que pueden ser estudiadas siguiendo un proceso similar al de este trabajo, tales como la variación porcentual diaria de la cotización, el volumen (en euros) de las transacciones de cada día o los valores máximos y mínimos de la jornada. En cuanto a las variables independientes, el modelo inicial utiliza once regresores: el número total de tuits que contienen al menos uno de estos términos, y el número de veces que se repite cada uno de ellos (Ucrania, Putin, Rusia, etc.) de manera individual.

Así, una vez definidas las variables, se procede a introducirlas en *Gretl* y a especificar que han de ser tratadas como una **serie temporal** con frecuencia diaria de 21 observaciones, iniciada el 18 de febrero y finalizada el 18 de marzo (el programa tiene incorporada una modalidad temporal que únicamente considera las fechas de lunes a viernes, lo que resulta idóneo para analizar valores bursátiles). Puesto que existe una hoja

de Excel para cada tipo de cotización (IBEX, Acciona, Acerinox, etc.), y que cada una de ellas presentará un modelo de regresión lineal diferente, el análisis debe ser repetido cada vez que se quiera analizar una nueva compañía. En este primer ejemplo analizaremos la hoja del IBEX 35 en su conjunto.

Tras incorporar los datos a *Gretl* y establecer la serie temporal, el siguiente paso consiste en construir un **modelo de mínimos cuadrados ordinarios** (MCO), en el que la variable dependiente es *Diferencialn* y los regresores son *TotalTuits*, *Ucrania*, *Putin*, *Rusia*, *Guerra*, *Invasión*, *Biden*, *Ejército*, *OTAN*, *Kiev* y *Crisis*. A continuación se obtiene un modelo como el siguiente:

Modelo 1: MCO, usando las observaciones 2022-02-18:2022-03-18 (T = 21)					
Variable dependiente: Diferencialn					
	<i>Coficiente</i>	<i>Desv. Típica</i>	<i>Estadístico t</i>	<i>Valor p</i>	
const	-0.0138830	0.00819031	-1.695	0.1209	
TotalTuits	0.000632544	0.000668936	0.9456	0.3666	
Ucrania	-0.00110613	0.000729832	-1.516	0.1606	
Putin	-0.00066818 2	0.000799698	-0.8355	0.4229	
Rusia	-0.00018435 1	0.000708366	-0.2602	0.8000	
Guerra	-0.00035950 1	0.000696199	-0.5164	0.6168	
Invasión	0.000216244	0.000705910	0.3063	0.7656	
Biden	-0.00019153 6	0.000835496	-0.2292	0.8233	
Ejercito	-0.00310941	0.00110360	-2.818	0.0182	**
OTAN	0.000666152	0.00175972	0.3786	0.7129	
Kiev	-0.00124944	0.000644467	-1.939	0.0813	*
Media de la vble. dep.	-0.001413		D.T. de la vble. dep.	0.022240	
Suma de cuad. residuos	0.002212		D.T. de la regresión	0.014873	
R-cuadrado	0.776375		R-cuadrado corregido	0.552750	
F(10, 10)	3.471777		Valor p (de F)	0.031169	
Log-verosimilitud	66.36465		Criterio de Akaike	-110.7293	
Criterio de Schwarz	-99.23955		Crit. de Hannan-Quinn	-108.2357	
rho	-0.111381		Durbin-Watson	2.100349	

Como se observa, únicamente existen dos variables (*Ejército* y *Kiev*) que cumplen con la condición de tener un **valor p** inferior a 0,10. Este requisito es necesario para garantizar que las variables sean **estadísticamente significativas**, es decir, que su

relación con la variable dependiente no sea producto del azar. Como el resto de regresores presentan valores demasiado altos, se procede a repetir el modelo suprimiendo *Biden*, que es la variable que presenta un valor p mayor (0,8233). El resultado es el siguiente:

Modelo 2: MCO, usando las observaciones 2022-02-18:2022-03-18 (T = 21)					
Variable dependiente: Diferencialn					
	<i>Coficiente</i>	<i>Desv. Típica</i>	<i>Estadístico t</i>	<i>valor p</i>	
const	-0.0138893	0.00782960	-1.774	0.1037	
TotalTuits	0.000512105	0.000395856	1.294	0.2223	
Ucrania	-0.00098698 7	0.000489844	-2.015	0.0690	*
Putin	-0.00054123 3	0.000551540	-0.9813	0.3475	
Rusia	-5.28988e- 05	0.000397605	-0.1330	0.8966	
Guerra	-0.00024437 2	0.000460939	-0.5302	0.6065	
Invasion	0.000301129	0.000574538	0.5241	0.6106	
Ejercito	-0.00295561	0.000837655	-3.528	0.0047	***
OTAN	0.000881533	0.00142241	0.6197	0.5480	
Kiev	-0.00114508	0.000436095	-2.626	0.0236	**

Media de la vble. dep.	-0.001413	D.T. de la vble. dep.	0.022240
Suma de cuad. residuos	0.002224	D.T. de la regresión	0.014218
R-cuadrado	0.775200	R-cuadrado corregido	0.591273
F(9, 11)	4.214709	Valor p (de F)	0.014180
Log-verosimilitud	66.30961	Criterio de Akaike	-112.6192
Criterio de Schwarz	-102.1740	Crit. de Hannan-Quinn	-110.3523
rho	-0.132984	Durbin-Watson	2.136807

Se puede observar ahora cómo el valor p de todas las variables ha disminuido, con la excepción de *Rusia*, que presenta un valor demasiado alto y que por ello será la siguiente variable eliminada. Se trata, pues, de repetir el modelo eliminando las variables que no resulten estadísticamente significativas, hasta llegar a uno que sí resulte interesante para el análisis. Además, al repetir los modelos, se debe mantener la atención en el valor **R-cuadrado**, ya que este valor nos indica cómo de fiable es el modelo, al reflejar qué proporción de la varianza de la variable está explicada por el modelo de regresión. Cuanto

más cercano se encuentre este valor a 1, mayor bondad del ajuste tendrá el modelo, y, por tanto más fiable, será.⁴

Así, los sucesivos modelos hasta llegar al definitivo presentan estos valores p:

Modelo	1	2	3	4	5	6	7
TotalTuits	0,37	0,22	0,0758	0,0137	0,0002	0,0002	$3,26 \times 10^{-5}$
Ucrania	0,16	0,06	0,040	0,0105	3×10^{-5}	4×10^{-5}	$2,96 \times 10^{-5}$
Putin	0,42	0,34	0,229	0,1190	0,0894	0,2079	-
Rusia	0,80	0,90	-	-	-	-	-
Guerra	0,62	0,61	0,466	0,2574	0,1881	-	-
Invasión	0,77	0,61	0,469	-	-	-	-
Biden	0,82	-	-	-	-	-	-
Ejército	0,018	0,005	0,003	0,0007	0,0003	0,0003	0,0003
OTAN	0,71	0,55	0,475	0,7337	-	-	-
Kiev	0,081	0,023	0,0026	0,0021	0,0015	0,0027	0,0014
R ²	0,776	0,775	0,774	0,7644	0,7622	0,7296	0,6984
R ² correg.	0,552	0,591	0,624	0,6375	0,6603	0,6395	0,6230

⁴ El valor R-cuadrado corregido especifica la proporción de las varianzas de la variable dependiente que tiene su explicación en los regresores, es decir, solo tiene en cuenta el impacto de los regresores que influyen en la variable dependiente.

Finalmente, para el caso del IBEX 35 en conjunto, hay cuatro variables estadísticamente significativas: *TotalTuits*, *Ucrania*, *Ejército* y *Kiev*. El modelo resultante sería el siguiente:

Modelo 7: MCO, usando las observaciones 2022-02-18:2022-03-18 (T = 21)					
Variable dependiente: Diferencialn					
	<i>Coficiente</i>	<i>Desv. Típica</i>	<i>Estadístico t</i>	<i>valor p</i>	
const	-0.0110766	0.00513280	-2.158	0.0465	**
TotalTuits	0.000422003	7.39882e-05	5.704	<0.0001	***
Ucrania	-0.00101644	0.000176621	-5.755	<0.0001	***
Ejercito	-0.00263200	0.000575521	-4.573	0.0003	***
Kiev	-0.00063873	0.000165260	-3.865	0.0014	***
4					

Media de la vble. dep.	-0.001413	D.T. de la vble. dep.	0.022240
Suma de cuad. residuos	0.002983	D.T. de la regresión	0.013655
R-cuadrado	0.698425	R-cuadrado corregido	0.623032
F(4, 16)	9.263718	Valor p (de F)	0.000451
Log-verosimilitud	63.22465	Criterio de Akaike	-116.4493
Criterio de Schwarz	-111.2267	Crit. de Hannan-Quinn	-115.3159
rho	0.093269	Durbin-Watson	1.744466

Este modelo demuestra que existe una **relación lineal** entre la variable *Diferencialn*, que refleja la diferencia entre la cotización del IBEX respecto al día anterior (ajustada por logaritmos naturales) y el número de veces que aparecen en Twitter los términos “Ucrania”, “Ejército” y “Kiev”, así como los diez términos del estudio en conjunto. Según este modelo, por ejemplo, un aumento de una unidad en el número de tuits que contengan la palabra “Ucrania” implicará una disminución del 0,1% en la cotización del IBEX, *ceteris paribus*. En cambio, un aumento de una unidad en el número de tuits que contengan alguna de las diez palabras analizadas aumentaría la cotización de la bolsa en un 0,04%. En conjunto, la ecuación lineal resultante de este modelo sería la siguiente:

$$\text{Diferencialn} = 0,000422 * \text{TotalTuits} - 0,001016 * \text{Ucrania} - 0,002632 * \text{Ejército} - 0,0006384 * \text{Kiev}$$

Este método presenta no obstante, un defecto notable, y es su **multicolinealidad**. Al desarrollar con *Gretl* este tipo de análisis, los resultados indican que para las variables

TotalTuits y *Ucrania* existe un alto grado de colinealidad (552 y 382, respectivamente), lo que indica que estas dos variables se comportan de manera bastante similar y que, por tanto, es complicado analizar sus efectos sobre la variable independiente de manera individual. Este problema se resolvería con una muestra más extensa o añadiendo variables al modelo, pero las limitaciones explicadas en el apartado anterior imposibilitan configurar un modelo que carezca de este defecto.

Así pues, una vez construido y depurado el modelo de regresión lineal múltiple que representa al IBEX de manera global, el resto del análisis consiste en **repetir** este proceso para cada una de las 35 empresas que lo componen, de manera **individual**. Los resultados de esta nueva fase, que difieren en mayor o menor medida de los obtenidos con el análisis general del IBEX en función de la compañía analizada, serán comentados en el siguiente apartado.

b. Resultados

Los resultados de cada una de las 35 empresas del IBEX presentan ciertas diferencias entre ellos, aunque presentan asimismo ciertos **patrones comunes**⁵. En primer lugar, se observan ciertas variables como *Biden* o *Invasión* que carecen de significancia estadística para la mayor parte de modelos, estando presentes únicamente en nueve y ocho de ellos, respectivamente. El caso más extremo es el de la variable *Crisis*, que es insignificante estadísticamente en todo los modelos. La razón más plausible de este fenómeno es la reducida frecuencia con que aparecen estos términos en Twitter en comparación con el resto de palabras seleccionadas.

Por otro lado, las variables *TotalTuits* y *Ejército* (ambas con significancia estadística en 22 modelos), y en menor medida *Ucrania* (21) y *Guerra* (19), están presentes en los modelos más precisos de la mayor parte de las compañías. Resulta interesante, asimismo, el caso de la variable *Kiev*, que pese a ser uno de los términos con menor número de menciones en Twitter, tiene significancia estadística en 18 de las 35 compañías del IBEX, así como en el modelo del IBEX global, desarrollado como ejemplo en la página anterior.

⁵ Para consultar el número de modelos en los que aparece cada término, ver Anexo 7.4.

Del mismo modo, se aprecia una notable diversidad en el **número de variables** estadísticamente significativas que componen cada modelo. En el caso de Telefónica, por ejemplo, hay hasta siete variables cuyo valor p es inferior a 0,10 y por tanto son significativas para el análisis. Además, tres de ellas presentan una correlación negativa con la variable dependiente (*Ucrania, Ejército y Kiev*), mientras que las otras cuatro están relacionadas de manera positiva (*Putin, Guerra, Invasión y OTAN*). Por último, el valor R-cuadrado es de 0,85, y el R-cuadrado corregido es de 0,77, lo que indica que es un modelo bastante fiable. En conclusión, se trata de un modelo con una utilidad bastante superior al del IBEX global y al de la mayoría de modelos individuales obtenidos en este análisis.

Por el contrario, para otras empresas como Acciona, los resultados son bastante menos reveladores. Para llegar a un modelo con más de una variable estadísticamente significativa es necesario eliminar todas las variables excepto *Guerra, OTAN y Kiev*, y aún en este caso, la variable *Guerra* presenta un valor p demasiado elevado (0,1057). Además, los valores R-cuadrado y R-cuadrado corregido son de 0,3636 y de 0,2395 respectivamente. Si se optara por eliminar la variable *Guerra*, estos valores disminuirían aún más (0,2426 y 0,1584), por lo que es preferible mantener esta variable, aunque se aleje ligeramente del objetivo de significancia estadística. En cualquier caso, ambos modelos presentan una utilidad bastante limitada, tanto por el número de variables que incluyen como por su bondad de ajuste.

Así pues, el análisis de regresión lineal múltiple desarrollado en este trabajo funciona mejor para unas compañías del IBEX que para otras, pero esto no quiere decir que sea un mal mecanismo de detectar correlación, sino más bien que las variables empleadas no afectan a las compañías de igual forma, o que se requiere un mayor número de regresores (o una serie temporal más larga) para poder construir modelos que presenten mayor robustez y utilidad. No obstante, en el caso de las compañías en las que se han podido determinar modelos de regresión valiosos, conviene conocer una utilidad adicional de estos: el análisis predictivo.

c. Predicciones

Como último aporte de este estudio de caso, conviene señalar las aplicaciones que este análisis econométrico puede tener en el área de la predicción. Si se dispone de

modelos sólidos y fiables, es posible pronosticar con mayor o menor precisión cómo afectará la **evolución futura** de las variables independientes a las variables dependientes. Evidentemente, el análisis histórico de cualquier índice bursátil demuestra que sus valores siguen generalmente un *random walk* (paseo aleatorio) imposible de predecir al 100%. No obstante, esa incertidumbre puede ser reducida en gran medida si se disponen de buenos **modelos predictivos**, lo que representa otra gran utilidad del uso del Big Data en el ámbito económico.

Para demostrar la importancia de esta aplicación, a continuación se desarrollará como **ejemplo** el caso de la empresa de energía Siemens Gamesa Renewable Energy. Con esta empresa se obtiene un modelo bastante sólido, que incluye ocho variables estadísticamente significativas (todas menos *Ejército*, *OTAN* y *Crisis*), y presenta unos valores R-cuadrado y R-cuadrado corregido de 0,8013 y 0,6688 respectivamente. Para comprobar las aplicaciones predictivas del modelo, se limita la serie temporal en *Gretl*, estableciendo un rango inferior al rango original (en este caso, se reduce la muestra de 21 a 18 días, hasta el 15 de marzo). Al efectuar estos cambios, el modelo varía ligeramente, y se hace necesario eliminar la variable *Biden* para mantener unos niveles de significancia estadística y de bondad de ajuste aceptables. El modelo resultante es el siguiente:

Modelo 2: MCO, usando las observaciones 2022-02-18:2022-03-15 (T = 18)
Variable dependiente: Diferencial

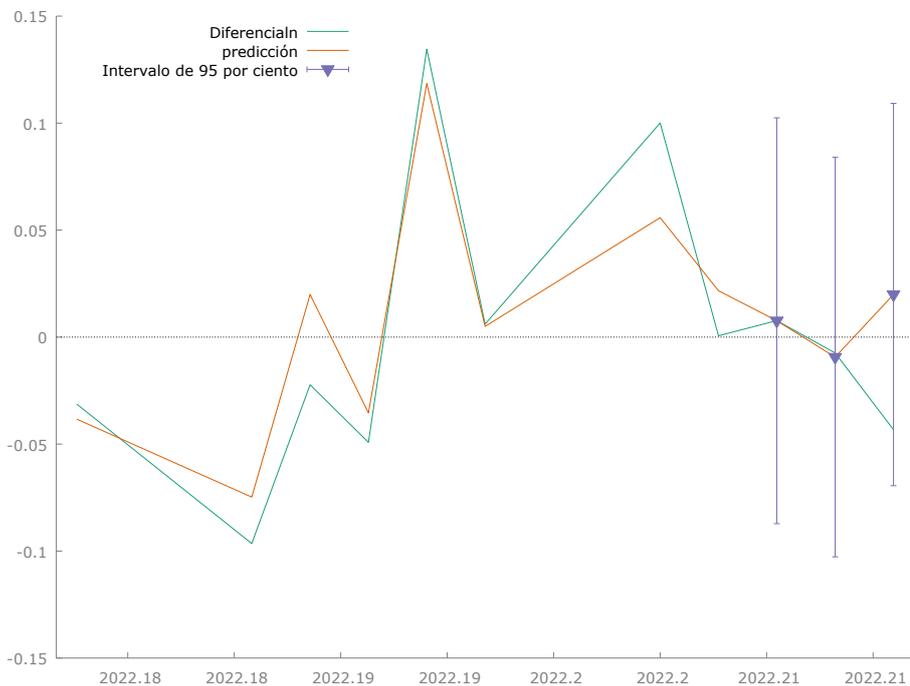
	<i>Coeficiente</i>	<i>Desv. Típica</i>	<i>Estadístico t</i>	<i>valor p</i>	
const	-0.0202943	0.0178490	-1.137	0.2820	
TotalTuits	0.00314026	0.000632374	4.966	0.0006	***
Ucrania	-0.00190026	0.000498582	-3.811	0.0034	***
Putin	-0.00566483	0.00125272	-4.522	0.0011	***
Rusia	-0.00375095	0.000980669	-3.825	0.0033	***
Guerra	-0.00439086	0.00100961	-4.349	0.0014	***
Invasion	-0.00202591	0.000924293	-2.192	0.0532	*
Kiev	-0.00603934	0.00107392	-5.624	0.0002	***

Media de la vble. dep.	0.006192		D.T. de la vble. dep.	0.057611
Suma de cuad. residuos	0.012926		D.T. de la regresión	0.035953
R-cuadrado	0.770915		R-cuadrado corregido	0.610555
F(7, 10)	4.807416		Valor p (de F)	0.013114
Log-verosimilitud	39.60916		Criterio de Akaike	-63.21831

Una vez ajustado el modelo, se procede a desarrollar un análisis predictivo hasta el 18 de marzo, es decir, hasta la fecha en que se dejan de obtener muestras de los regresores (número de tuits). El resultado, para un intervalo de confianza del 95%, es el siguiente:

Para intervalos de confianza 95%, $t(10, 0.025) = 2.228$

Observaciones	Diferencial	predicción	Desv. típica	Intervalo de 95%
2022-03-16	0.00770539	0.00764882	0.0425595	(-0.0871796, 0.102477)
2022-03-17	-0.00739831	-0.00933396	0.0418978	(-0.102688, 0.0840201)
2022-03-18	-0.0432760	0.0198869	0.0400907	(-0.0694408, 0.109215)



Como se puede observar en la tabla y en el gráfico, el modelo predice con bastante **exactitud** la evolución de la cotización de Siemens Gamesa durante los días 16 y 17 de marzo, si bien el último día la estimación se aleja considerablemente del valor observado en el escenario real. Esta imprecisión, no obstante, tiene sentido desde un punto de vista

estadístico, puesto que el modelo construido con *Gretl* dista en gran medida de ser perfecto (su R-cuadrado corregido es solo de 0,6106) a causa de las limitaciones anteriormente comentadas, y como se ha señalado anteriormente, los valores bursátiles siguen un patrón extremadamente difícil de predecir, aun contando con elevados niveles de información.

En todo caso, perfeccionando los modelos a través de series temporales más extensas y variables más significativas, es posible llevar a cabo análisis predictivos con un índice de acierto mucho mayor, lo cual resulta de gran utilidad para estudiar posibles escenarios en los que las variables independientes actúen de una u otra forma, y pronosticar cómo estas evoluciones pueden afectar a la variable dependiente. Se trata, una vez más, de convertir un enorme volumen de datos en **información de calidad** de una manera extremadamente rápida y sencilla, siendo esta la gran utilidad del análisis del Big Data.

5. Conclusiones

A través del estudio de caso desarrollado en este trabajo, ha sido posible comprobar empíricamente la importancia que tiene hoy en día el Big Data en distintos ámbitos de la valoración. En concreto, este trabajo se ha centrado en el valor de las acciones de las 35 compañías que componen el IBEX 35, lo que compone un elemento esencial de la valoración según métodos como los basados en el PER (*Price to Earnings Ratio*) y otros múltiplos como el que relaciona la cotización de las acciones de una compañía con su valor contable.

La primera conclusión que se puede sustraer de este trabajo es la importancia de contar con herramientas adecuadas para poder extraer información de calidad del Big Data. Los programas empleados durante el estudio de caso (Excel, *Vicinitas* y *Gretl*) permiten ejecutar tareas complejas que engloban miles de datos de manera prácticamente automática, y ofrecen numerosas aplicaciones adicionales que no han llegado a ser exploradas en este trabajo. Así pues, resulta imprescindible disponer de estas herramientas y contar con equipos que sepan manejarlas correctamente para poder extraer la máxima información posible y desarrollar de este modo valoraciones de la mayor calidad y profundidad.

Un segundo aspecto fundamental para este tipo de análisis es la calidad de los datos utilizados. Como se ha explicado en el apartado *Datos*, la planificación inicial de este trabajo contemplaba el tratamiento de millones de datos, al relacionar la cotización del IBEX 35 durante varios meses con todos los tuits generados durante ese período de tiempo. Además, el plan inicial contemplaba utilizar como fuente de datos la totalidad de la plataforma Twitter en español, y no únicamente los once periódicos más relevantes del panorama financiero y generalista de España. Sin embargo, estas limitaciones pueden ser superadas con una mayor inversión de tiempo y dinero, puesto que existen herramientas más sofisticadas que son capaces de sustraer datos durante un periodo de tiempo continuo y pueden generar así líneas temporales de una mayor longitud.

Así pues, el primer objetivo de este trabajo, que consistía en demostrar la utilidad del Big Data y de sus herramientas en el área de la valoración de compañías, ha sido cumplido con creces. A pesar de contar con numerosas limitaciones, el estudio de caso ha logrado construir 35 modelos de regresión lineal múltiple que permiten estudiar con distintos grados de fiabilidad la relación entre la cotización de cada empresa del IBEX 35 y la

aparición de determinados términos en las redes sociales. Evidentemente, esta relación se trata de un mero ejemplo, puesto que la evolución del mercado bursátil depende de una infinidad de factores, y el conflicto entre Rusia y Ucrania es un evento internacional puntual cuya duración no es probable que se prolongue demasiado tiempo (y de hacerlo, los tuits relacionados con el conflicto serían cada vez menos relevantes).

No obstante, a pesar de tratarse de un ejemplo con una aplicación bastante limitada, no deja de ser por ello un buen reflejo de cómo con las herramientas adecuadas se puede obtener información de manera relativamente fácil y sencilla. Si, por ejemplo, se pretende conocer cómo afectaría un posible ataque ruso sobre Kiev al valor de mercado de Telefónica, resulta bastante útil disponer de un modelo que prediga que por cada mención a la capital ucraniana en Twitter España, el valor de la acción de esta compañía se reduce en un 0,06%, mientras que por cada mención a la OTAN, este valor aumenta en un 0,19% (manteniendo constantes el resto de variables).

En cuanto al segundo objetivo, orientado a desarrollar los ámbitos del Big Data y de la valoración de empresas, así como la combinación de ambos, este trabajo describe y sintetiza las principales dimensiones y aplicaciones prácticas de ambos ámbitos, reflejando finalmente la utilidad que puede tener combinar ambas áreas a través del ejemplo desarrollado en el análisis práctico. No obstante, en el caso del Big Data, es necesario puntualizar que se trata de un ámbito tan amplio y novedoso que existe un gran número de aplicaciones y herramientas en proceso de desarrollo o implementación, algunas de las cuales resultarán imprescindibles en unos años y convertirán en obsoletas a las empleadas mayoritariamente hoy en día. Por lo tanto, es conveniente mantenerse actualizado acerca de las distintas posibilidades que ofrece el Big Data en prácticamente cualquier ámbito de análisis, especialmente si se pretende obtener información de calidad que otorgue una posición comparativa ventajosa en un determinado sector o mercado.

Por último, el tercer objetivo de este trabajo consistía en demostrar la importancia que tienen actualmente las redes sociales y lo que en ellas se publica para los mercados financieros. Numerosos estudios y teorías financieras han analizado la relación entre el grado de información que existe en un determinado mercado y sus niveles de eficiencia y flexibilidad. Sin embargo, desde la aparición de las redes sociales, la información se transmite con una velocidad nunca antes conocida, y las fuentes de información son bastante más numerosas y diversas ahora en comparación con las de unos años atrás. Si bien esta nueva realidad ha dificultado sustancialmente el análisis de la información que

afecta a los mercados financieros, ofrece al mismo tiempo nuevas posibilidades de tratar de comprender y predecir determinados patrones gracias al uso de nuevas fuentes de datos y herramientas de análisis.

Así pues, en conclusión, este trabajo ha cumplido con sus tres objetivos, y a través de un ejemplo concreto, ha demostrado la importancia que tiene el Big Data hoy en día en el proceso de valoración de compañías, independientemente del método de valoración empleado o del propósito de esta valoración. Como se explica en el marco conceptual, son tantas y tan variadas las aplicaciones que ofrecen actualmente este tipo de herramientas, que cualquier persona o entidad interesada en analizar grandes volúmenes de datos puede acceder a una gran cantidad de información de calidad con una facilidad impensable hace tan solo unas décadas.

No obstante, no es esta evolución hasta el día de hoy el aspecto más prometedor del Big Data, sino las posibilidades que ofrecerá en el futuro y que aún quedan por explorar. Al ritmo al que avanza la tecnología, resulta difícil imaginar el volumen de datos que seremos capaces de procesar en unos años, y las oportunidades, aunque también amenazas, que ello conllevará. Por el momento, estas posibilidades abarcan todo tipo de ámbitos, y están al alcance de cualquiera que disponga de tres elementos: datos de calidad, herramientas capaces de procesarlos e interés por ir más allá de las cifras y obtener verdadera información.

6. Bibliografía

Abellán, J. L. (s.f.). *CAPEX*. Economipedia. Recuperado el 10 de febrero de: <https://economipedia.com/definiciones/capex.html>

Arias, A.S. (s.f.). *PER ratio*. Economipedia. Recuperado el 9 de febrero de 2022 de: <https://economipedia.com/definiciones/per-ratio.html>

Balode, N. (4 de marzo de 2022). *Top 5 social media scraping tools for 2022*. Dripify. Recuperado el 10 de marzo de 2022 de: <https://dripify.io/top-5-social-media-scraping-tools-for-2020-2021/>

Barrios, A. P. (2013). Valoración de empresas: Métodos de valoración. *Contexto*, 2(1), 84-100.

Bonilla, F. L. (2010). El valor económico agregado (EVA) en el valor del negocio. *Revista Nacional de administración*, 1(1), 55-70.

Cárdenas, M. V., y Morales, J. A. R. (2016). Evaluación de la gestión financiera usando variables latentes en modelos estocásticos de fronteras eficientes. *Dyna*, 83(199), 35-40.

Casagrande, F. (2021). *Big Data Valuation. The role of data in corporate valuation*. [Trabajo de Fin de Máster, Università Ca'Foscari Venezia].

Comisión Europea. (s.f.). *Protección de datos. Normas sobre protección de datos personales dentro y fuera de la UE*. Recuperado el 12 de febrero de 2022 de: https://ec.europa.eu/info/law/law-topic/data-protection_es

Deloitte (2013). *Corporate Development 2013. Pushing Boundaries in M&A*. Recuperado el 7 de febrero de 2022 de: <https://www2.deloitte.com/za/en/pages/mergers-and-acquisitions/articles/corp-dev-2013.html>

Erasmus, P.D., y Lambrechts, I. J. (2006). EVA and CFROI: A comparative analysis. *Management Dynamics: Journal of the Southern African Institute for Management Scientists*, 15(1), 14-26.

Esteban, M.V., Moral, M. P., Orbe. S., Regúlez, M., Zarraga, A., y Zubia, M. (2009). *Análisis de regresión con Gretl*. Universidad del País Vasco.

Fanning, K., y Drog, E. (2014). Big Data: New opportunities for M&A. *Journal of Corporate Accounting & Finance*, 25(2), 27-34.

Fernández, P. (2008). *Métodos de valoración de empresas*. IESE Business School – Universidad de Navarra.

Fernández, P. (2008a). *Valoración de empresas por descuento de flujos: 10 métodos y 7 teorías*. IESE Business School – Universidad de Navarra.

García, Y. (14 de octubre de 2020). *¿Qué es ARPU? Conoce cómo funciona y sus beneficios*. IEBS. Recuperado el 10 de febrero de 2022 de: <https://www.iebschool.com/blog/que-es-arpu-conoce-como-funciona-y-sus-beneficios-marketing-digital/>

Granados, R. M. (2016). *Modelos de regresión lineal múltiple*. Universidad de Granada.

Hurtado De Mendoza, J. M. (2020). *Adquisición, normalización, enriquecimiento y visualización de eventos de ciberseguridad con tecnologías Big Data*. [Trabajo de Fin de Grado, Universidad de La Laguna].

Instituto de Ingeniería del Conocimiento. (s.f.). *7 herramientas Big Data para tu empresa*. Recuperado el 1 de marzo de 2022 de: <https://www.iic.uam.es/innovacion/herramientas-big-data-para-empresa/>

IVSC (31 de enero de 2022). *International Valuation Standards 2022*. Recuperado el 9 de febrero de 2022 de: https://www.appraisers.org/docs/default-source/default-document-library/ivsc-effective-31-jan-2022.pdf?sfvrsn=16c86ad4_2

López, J. F. (s.f.). *Coefficiente de determinación (R cuadrado)*. Economipedia. Recuperado el 10 de marzo de: <https://economipedia.com/definiciones/r-cuadrado-coeficiente-determinacion.html>

Martins, A., Martins, P., Caldeira, F., y Sá, F. (2020). An Evaluation of How Big-Data and Data Warehouses Improve Business Intelligence Decision Making. En Rocha, Á., *Trends and Innovations in Information Systems and Technologies*, (609–619). Springer.

Microsoft Dynamics. (s.f.). *ERP definition*. Microsoft. Recuperado el 11 de febrero de 2022 de: <https://dynamics.microsoft.com/es-es/erp/define-erp/>

Moriarty, R., Ly, H., Lan, E., y McIntosh, S. K. (Diciembre, 2019). Deal or No Deal: Predicting Mergers and Acquisitions at Scale. *2019 IEEE International Conference on Big Data (Big Data)*, 5552-5558.

Moro Visconti, R., Montesi, G., y Papiro, G. (Primavera de 2018). Big data-driven stochastic business planning and corporate valuation. *Corporate Ownership & Control*, 15(3-1), 189-204

Oliete, B. (27 de marzo de 2012). *Análisis y gestión de las Necesidades Operativas de Fondos (NOF)*. EmpresaActual.com. Recuperado el 10 de febrero de: <https://www.empresaactual.com/las-necesidades-operativas-de-fondos/>

ONEtoONE Corporate Finance. (s.f.). *Big Data y su futuro gran impacto en M&A*. Recuperado el 8 de febrero de 2022 de: <https://www.onetonecf.com/es/big-data-en-ma/>

Otto, C. (9 de octubre de 2016). ¿Es legal que Facebook y Whatsapp compartan tus datos? *La Vanguardia*. Recuperado el 3 de febrero de 2022 de: <https://www.lavanguardia.com/tecnologia/20161009/41843913284/facebook-whatsapp-compartan-datos.html>

Pariante, R., y Reig, P. (10 de septiembre de 2018). ¿Qué es el fondo de comercio? *BBVA*. Recuperado el 10 de febrero de 2022 de: <https://www.bbva.com/es/que-es-el-fondo-de-comercio/>

Pérez, A. (18 de marzo de 2021). Outsourcing Data Science. *Audalia Nexia*. Recuperado el 3 de febrero de 2022 de: https://www.audalianexia.com/blog/advisory/outsourcing_data

PWC. (2019). *Putting a value on data*. Recuperado el 2 de febrero de 2022 de: <https://www.pwc.co.uk/issues/data-analytics/insights/putting-value-on-data.html>

Rankia (27 de diciembre de 2019). *¿Qué es el R cuadrado ajustado?* Recuperado el 10 de marzo de 2022 de: <https://www.rankia.mx/blog/como-comenzar-invertir-bolsa/4439147-que-r-cuadrado-ajustado>

Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE (Reglamento general de protección de datos). Recuperado el 10 de febrero de 2022 de: <https://eur-lex.europa.eu/legal-content/es/ALL/?uri=CELEX:32016R0679>

Rodó, P. (s.f.). *Logaritmos en econometría*. Economipedia. Recuperado el 8 de marzo de: <https://economipedia.com/definiciones/logaritmos-en-econometria.html>

Salesforce (s.f.). *¿Qué es CRM?* Salesforce. Recuperado el 8 de febrero de: <https://www.salesforce.com/mx/crm/>

Selby, J. (20 de abril de 2015). The impact of Big Data decisions on Business Valuations. *Datafloq*. Recuperado el 31 de enero de 2022 de: <https://datafloq.com/read/impact-big-data-decisions-business-valuation/1041>

Stander, J. B. (2015). *The modern asset: big data and information valuation*. [Tesis doctoral, Stellenbosch University].

Statista. (Marzo de 2018). *Big data market size revenue forecast worldwide from 2011 to 2027*. Recuperado el 2 de febrero de 2022 de: <https://www.statista.com/statistics/254266/global-big-data-market-forecast/>

Syed, A., Gillela, K., y Venugopal, C. (2013). The future revolution on big data. *Future*, 2(6), 2446-2451.

Ucha, A. P. (s.f.). *WACC – Coste medio ponderado del capital*. Economipedia. Recuperado el 10 de febrero de: <https://economipedia.com/definiciones/coste-medio-ponderado-del-capital-wacc.html>

Universidad de Alcalá (3 de septiembre de 2021). *Método EVA como valoración de finanzas cuantitativas*. Recuperado el 4 de febrero de 2022 de: <https://www.master-finanzas-cuantitativas.com/eva-master-en-finanzas/>

Valls, E. (s.f.). *Deuda Financiera Neta*. Rankia. Recuperado el 10 de febrero de 2022 de: <https://www.rankia.com/diccionario/economia/deuda-financiera-neta>

Wiener, M., Saunders, C., & Marabelli, M. (2020). Big-data business models: A critical literature review and multiperspective research framework. *Journal of Information Technology (Sage Publications Inc.)*, 35(1), 66–91. Recuperado el de 8 de febrero de 2022 de: <https://doi.org/10.1177/0268396219896811>

Wise, J. (8 de marzo de 2022). *11 best social media scraping tools for data extraction in 2022*. Earthweb. Recuperado el 10 de marzo de 2022 de: <https://earthweb.com/social-media-scraping-tools/>

Wise, J. (8 de marzo de 2022a). *7 best Twitter scrapers for scraping data & tweets in 2022*. Earthweb. Recuperado el 10 de marzo de 2022 de: <https://earthweb.com/twitter-scrapers/>

Zhang, L. (2020). M&A Financial Performance Analysis. *The 4th International Conference on Business and Information Management (ICBIM)*, 8-15. https://dl.acm.org/doi/abs/10.1145/3418653.3418670?casa_token=11rGDHxSz1kAAAAA:FrEl6ym6QhWAqT3TYI4NFh_7i5ZSccqwS8ILfU_MVNTL7s2EOwCwhJqzH0CZ3DyONZMgGqGnEPp3

7. Anexos

1. Índice de términos

- ARPU: *Average Revenue Per User* (ingresos promedio por usuario). Valor que refleja los ingresos totales que una empresa espera generar por cada usuario nuevo capturado.
- BDBM: Siglas de *Big Data Business Models*, es decir, modelos de negocio en los que el uso del Big Data supone la actividad principal de la compañía.
- Big Data: Agrupaciones de datos que integran piezas de información diversas y complejas y que requieren de herramientas especiales para poder ser procesados.
- CAPEX: *Capital Expenditure* (gasto en capital). Inversión en capital o inmovilizado fijo que realiza una compañía para incorporar o mejorar su activo no corriente.
- Ceteris paribus: Expresión latina que significa “manteniéndose el resto de elementos constantes”.
- CFROI: *Cash flow return on investment*. Valor combinado de todos los flujos de caja que una compañía genera en un periodo determinado expresado como porcentaje del efectivo invertido en los activos de esa compañía.
- CRM: Siglas de *Customer Relationship Management*. Conjunto de procesos utilizados por una compañía para gestionar y analizar las interacciones con sus clientes.
- Data set: Conjunto de datos con unas características o un propósito en común.
- Deuda Financiera Neta: Resultado de restar la deuda financiera de una compañía menos los activos financieros corrientes y el efectivo.
- EBIT / BAIT: Beneficio antes de intereses e impuestos. Cifra de la Cuenta de Resultados de una compañía que resulta de la sustraer los costes de explotación y demás gastos generales de la cifra de ventas.
- EBITDA: Beneficio antes de amortización, intereses e impuestos. Cifra de la Cuenta de Resultados de una compañía que resulta de la sustraer la amortización al EBIT.
- EVA: *Economic Value Added* (Valor Económico Agregado). Método de valoración de empresas que cuantifica la generación de valor que ha tenido lugar en una compañía concreta durante un período de tiempo determinado.
- ERP: *Enterprise Resource Planning* (Planificación de Recursos de la Empresa). Sistemas que ayuda a automatizar y gestionar procesos de utilización de los recursos de una compañía (financieros, humanos, operacionales...).

- GDPR / RGPD: Siglas de *General Data Protection Regulation*. Reglamento de la Unión Europea de 2016 relativo a la protección de las personas físicas en lo referente al tratamiento de sus datos personales y la libre circulación de estos.
- M&A: *Mergers & Acquisitions* (Fusiones y Adquisiciones). Bajo este nombre se conoce a los procesos por los que dos o más compañías se unen originando una nueva.
- NOF: Necesidades Operativas de Fondos. Inversiones netas que debe realizar una compañía en operaciones corrientes de su actividad.
- PER: *Price to Earnings Ratio* (ratio precio-beneficio). Ratio financiera que pone en relación el precio de la acción de una compañía con su beneficio por acción (BPA).
- WACC: *Weighted Average Cost of Capital* (coste medio ponderado del capital). Coste de los recursos propios de una compañía ponderando el tamaño relativo de la deuda financiera y los fondos propios.
- *Value Drivers*: Elementos de una compañía que fomentan la generación de valor.

2. Tabla de periódicos más seguidos en Twitter España

Periódico	Seguidores (miles)
El País	8400
El Mundo	4100
ABC	2100
La Vanguardia	1200
elDiario.es	1300
El Confidencial	958
Expansión	750
ElEconomista	719
CincoDias	349
BolsaManía	59
Cotizalia	31

3. Tabla de tuits por día y hora

	17-mar	16-mar	15-mar	14-mar	11-mar	10-mar	09-mar	08-mar	07-mar	04-mar	03-mar
Ucrania	163	202	206	215	239	250	218	241	306	850	307
Putin	70	69	58	54	87	72	73	57	73	284	104
Rusia	89	125	118	133	142	168	152	161	183	458	191
Guerra	106	139	127	133	110	132	103	121	140	485	147
Invasión	33	31	28	31	42	32	24	49	63	191	48
Biden	17	25	20	0	15	7	15	27	1	29	4
Ejército	7	10	6	7	13	12	3	13	13	73	15
OTAN	15	26	23	19	18	15	14	21	13	83	6
Kiev	17	37	56	19	15	20	17	30	28	148	9
Crisis	14	24	28	22	11	8	22	21	15	53	7
TOTAL	531	688	670	633	692	716	641	741	835	2654	838

	02-mar	01-mar	28-feb	25-feb	24-feb	23-feb	22-feb	21-feb	18-feb	TOTAL
Ucrania	308	1404	852	709	140	35	39	21	308	7160
Putin	100	412	296	287	39	3	25	13	100	2335
Rusia	169	751	476	466	87	25	31	9	169	4191
Guerra	150	678	445	370	40	6	4	8	150	3679
Invasión	64	204	114	136	18	6	3	1	64	1215
Biden	22	56	28	21	12	3	5	4	22	353
Ejército	11	48	46	54	5	0	0	2	11	358
OTAN	12	87	48	85	8	1	0	1	12	520
Kiev	19	165	14	98	10	1	2	3	19	741
Crisis	14	77	66	79	11	4	17	6	14	529
TOTAL	869	3882	2385	2305	370	84	126	68	869	21081

4. Presencia de cada variable en cada modelo

Mod.	TotalTuits.	Ucrania	Putin	Rusia	Guerra	Invas.	Biden	Ejércit	OTAN	Kiev	Crisis
Acciona					X				X	X	
Acerinox		X	X		X			X			
ACS	X			X			X		X		
Aena								X	X	X	
Almirall	X	X	X						X		
Amadeus	X	X						X		X	
ArcelorMittall		X	X	X				X			
Banco Sabadell	X		X	X	X	X	X	X	X		
Bankinter	X			X	X		X		X		
BBVA		X		X	X		X	X			
CaixaBank		X		X	X		X	X			
Cellnex	X			X			X				
Cie Automotive	X	X						X		X	
Colonial	X		X	X		X		X	X		
Enagás	X	X	X		X					X	
Endesa	X	X								X	
Ferrovial								X	X	X	
Fluidra				X	X			X	X	X	
Grifols	X	X	X	X	X		X	X		X	
IAG	X	X						X		X	
Iberdrola	X	X	X	X	X		X	X		X	
Inditex	X	X	X		X			X		X	
Indra		X				X			X		
Mapfre		X		X	X	X		X			
Meliá Hotels	X		X					X	X		
Merlin Properties	X		X	X	X	X		X	X		
Naturgy Energy	X	X	X		X					X	
PharmaMar						X		X	X	X	
Red Eléctrica	X								X	X	
Repsol*											
Rovi	X	X		X	X			X			
Santander	X	X		X	X			X			
Siemens Gamesa	X	X	X	X	X	X	X			X	
Solaria	X	X	X	X	X					X	
Telefónica		X	X		X	X		X	X	X	
TOTAL	22	21	15	17	19	8	9	22	15	18	0

* Para la compañía Repsol no ha sido posible construir un modelo con variables estadísticamente significativas cuyo R-cuadrado corregido tuviera un valor positivo.