



**GRADO EN ADE Y GRADO EN DERECHO**

**FACULTAD DE CIENCIAS  
ECONÓMICAS Y EMPRESARIALES**

**TRABAJO FIN DE GRADO  
CURSO ACADÉMICO [2021-2022]**

**TÍTULO:**

**La generación de ventaja competitiva en PYMES a través de big data y la filosofía de gestión data-driven**

**AUTOR:**

**Pablo Aguilar Borrero**

**TUTOR:**

**D. Carlos Arias Martín**

**DEPARTAMENTO:**

**DEPARTAMENTO DE ECONOMÍA APLICADA I**

**ÁREA DE CONOCIMIENTO:**

**MÉTODOS CUANTITATIVOS PARA LA ECONOMÍA Y LA EMPRESA**

**RESUMEN:**

El auge de las tecnologías de análisis con herramientas de Big Data, supone un cambio en la gestión empresarial, dedicándose este estudio a analizar las causas de tal transformación, la mecánica básica de funcionamiento de estas herramientas y su incidencia en la aparición de nuevos modelos de gestión, así como su potencial para la generación de ventaja competitiva en la pequeña y mediana empresa.

**PALABRAS CLAVE:**

Big Data; data-driven; ventaja competitiva; cultura empresarial; tecnología.

# ÍNDICE

<b>1. Capítulo 1: Introducción.....</b>	<b>5</b>
1.1. Surgimiento y evolución del big data.....	5
1.2. Concepto de big data, utilidades y aplicaciones generales.....	5
1.3. Herramientas asociadas al big data.....	6
1.3.1 Business intelligence.....	6
1.3.2 Machine learning.....	6
1.4. Las 5 V'S del big data.....	7
<b>2. CAPÍTULO 2: Objetivos del trabajo de investigación.....</b>	<b>9</b>
2.1. Objetivos generales: el análisis de la generación de ventaja competitiva mediante el uso de big data en el entorno corporativo actual.....	9
2.2. Objetivos específicos.....	9
<b>3. CAPÍTULO 3: Uso y aplicación del big data en el entorno corporativo actual.....</b>	<b>11</b>
3.1. Ámbito de aplicación de big data en base a actividad y tamaño organizativo.....	11
3.2. Tipos de datos válidos para big data.....	14
3.3. Metodología, tratamiento y herramientas de recolección de datos.....	14
3.4. Limitaciones operativas del big data.....	16
<b>4. CAPÍTULO 4: Regulación y paradigma legal de big data a nivel nacional y europeo.....</b>	<b>17</b>
4.1. Regulación del big data en el ámbito de la Unión Europea.....	17
4.2. Legislación nacional, directiva 95/46/CE y Reglamento 2016/679.....	18
4.3. Incidencia del paradigma legal en el potencial de big data.....	19
<b>5. CAPÍTULO 5: La utilidad empresarial de big data.....</b>	<b>21</b>
5.1. Organización dirigida por datos.....	21
5.2. ¿Qué caracteriza a la organización data driven?.....	22
5.3. ¿Qué ventaja competitiva aporta ser una organización data driven?.....	22
5.4. ¿Cómo transformarnos en una organización basada en datos?.....	23
<b>6. CAPÍTULO 6: La obtención de ventaja competitiva en la organización basada en datos.....</b>	<b>25</b>
6.1. Ganancias de eficiencia e innovación.....	25
6.2. Conceptos técnicos alrededor de big data.....	27
6.3. Map Reduce, Hadoop y Spark.....	28
6.4. La adopción del análisis de macrodatos y la filosofía data driven en pymes.....	28

<b>7. CAPÍTULO 7: Big data, ética y responsabilidad social corporativa.....</b>	<b>31</b>
7.1.El futuro para big data.....	31
7.2.Ética,responsabilidad y propuesta regulatoria.....	31
<b>8.CAPÍTULO 8:CONCLUSIONES FINALES.....</b>	<b>33</b>
<b>9.CAPÍTULO 9:BIBLIOGRAFÍA.....</b>	<b>33</b>

## **CAPÍTULO 1 INTRODUCCIÓN**

El siguiente estudio tratará de analizar e identificar las posibilidades para la generación de ventaja competitiva en pequeñas y medianas empresas en atención a las particularidades de su entorno respecto a las grandes corporaciones. Para el desarrollo de la tarea propuesta se partirá de un concepto y aplicaciones generales de las herramientas de “*big data*” o análisis de datos masivos y sus asociadas, para progresivamente analizar su utilización y grado de implantación en grandes corporaciones hasta concluir dicho análisis en las particularidades del entorno PYME nacional y en un modelo básico de implantación para estas.

La elección de la temática de este trabajo tiene su origen en la creciente presencia en la vida cotidiana de las herramientas de análisis de datos masivos y su importancia también creciente en la realidad digital y la expansión de esta como consecuencia del COV SARS-2.

### **1.1 SURGIMIENTO Y EVOLUCIÓN DEL BIG DATA**

La historia del “*big data*” no tiene un momento concreto de nacimiento, sino un período de evolución en forma de progreso tecnológico, sin embargo el término es acuñado por primera vez en 1989 por Erik Larson para referirse al uso de datos de clientes en el marketing. Sin embargo el “*big data*” como lo conocemos hoy día comienza a configurarse en 1997, momento en el que Google lanza su motor de búsqueda y con él comienza una generación de datos respecto a sus usuarios sin precedentes, pero estas nuevas herramientas de análisis de datos alcanzan su concepción actual a partir del año 2005, con el nacimiento del “internet social”, permitiendo elaborar verdaderos perfiles individuales que permitían modelizar rasgos de la población en todos los aspectos de su vida en base a la información compartida en la red o el uso de esta que hacían.

### **1.2 CONCEPTO DE BIG DATA, UTILIDADES Y APLICACIONES GENERALES.**

El big data o los macrodatos hacen referencia a un conjunto de datos de gran tamaño y complejidad, que ya sean de carácter cuantitativo o cualitativo, permiten mediante el uso de softwares especializados la modelización y recogida de patrones que arrojan información más allá de la inferencia estadística tradicional. Si bien los usos de estos macrodatos deben sus inicios a aplicaciones comerciales, en la actualidad podemos ver una extensión a cualquier rama de la actividad humana en la que se generen grandes muestras de datos a partir de las que inferir un modelo altamente complejo que permita predecir el comportamiento poblacional.

(Mayer-Schönberger y Cukier,2016)

Es una realidad que el big data encuentra su máxima expresión en las funciones comerciales empresariales y en especial en la segmentación de clientes. Sin embargo el ejemplo más reciente la expansión a otros ámbitos como la salud pública la encontramos en el uso del análisis de big data para el análisis, control y toma de decisiones en cuanto a medidas restrictivas durante la pandemia del COVID-19 y a su vez en la toma de decisiones relativas a la producción de vacunas y a la forma de realizar las campañas de vacunación, lo que nos devuelve al ámbito empresarial a través de una modelización de la demanda y distribución del fármaco.

(Alsunaidi SJ; Almuhaideb AM;Ibrahim NM; Shaikh FS; Alqudaihi KS; Alhaidari FA,;Khan IU, Aslam N y Alshahrani MS,2021)

Por tanto y a modo de conclusión respecto a las utilidades y aplicaciones de los macrodatos actualmente podemos ver una expansión desde sus orígenes puramente comerciales a cualquier ámbito de la actividad humana en el que se requiera de la toma de decisiones, estando éstas sujetas a un número elevado de variables.

### **1.3 HERRAMIENTAS ASOCIADAS AL “BIG DATA”**

El “*big data*” se centra en la captura, almacenamiento y procesamiento de los datos , sin embargo la utilidad de esto es limitada en su uso individual , es por ello que nos encontramos con una serie de técnicas y herramientas que partiendo de estos datos y de los modelos generados pretenden generar utilidad. Al nivel empresarial que nos ocupa nos centraremos para nuestro análisis fundamentalmente en dos: “*business intelligence*” y “*machine learning*”.

#### **1.3.1 “Business Intelligence”**

El “*business intelligence*” es un concepto referido al conjunto de técnicas y procesos de análisis de dichos datos obtenido por “*big data*” y su transformación en una información cuyo formato y contenido permita la toma de decisiones de gestión a todos los niveles corporativos. Si bien “*big data*” y “*business intelligence*” forman parte de una misma realidad, el “*business intelligence*” busca ofrecer una visión integrada y global de los propios macrodatos sobre la realidad del entorno empresarial y expresarlos en un formato útil para la toma de decisiones, a fin de generar ventaja competitiva. (Cano,2007)

Por ende, podemos distinguir que mientras el concepto estricto de “*big data*” solo se centra en unos datos de carácter masivo. El concepto de “*business intelligence*” se centra en qué hacer con esos datos a fin de generar ventaja competitiva a partir de la aplicación de estos en una modelización de negocio.

#### **1.3.2 “Machine Learning”**

El “*machine learning*” es una de las máximas expresiones del uso de los macrodatos en nuestra realidad, está presente en las recomendaciones de las plataformas de ocio digital, la publicidad en la red, la detección del blanqueo de capitales o el diagnóstico médico. Se trata de una rama de la inteligencia artificial que permite a las máquinas aprender sin necesidad de ser constantemente reprogramadas, sino en base a los macrodatos y al reconocimiento de patrones dentro de estos. El funcionamiento básico del “*machine learning*” se basa en un constante análisis de datos para obtener el resultado óptimo, siendo su única base el Teorema de Bayes combinado con la computación moderna para generar redes neuronales como modelo computacional que permite no solo el almacenado y procesado de la información, sino también su interpretación en términos similares a los del cerebro humano. (Burrel,2016)

Por tanto y comparación con los propios “*big data*” y “*business intelligence*”, el “*machine learning*” es el siguiente nivel en la utilización de macrodatos, no solo los capta, gestiona y modeliza; sino que además los analiza e interpreta por sí mismo a la semejanza del cerebro humano mediante el uso del Teorema de Bayes, publicado en 1763, pero instrumentalizado por la computación moderna para construir un modelo de pensamiento y aprendizaje puramente estadístico. Es por ello que el “*machine learning*” elimina el componente emocional de las decisiones humanas, refinando dichas decisiones a través del uso de datos objetivos.

#### 1.4 LAS 5 V's DEL “BIG DATA”

“*Big data*” es un concepto enormemente abstracto, puesto que en base a las definiciones expuestas cualquier dato que expresa una característica dentro de una población subjetivamente considerada amplia podría encuadrarse en el concepto de macrodato.

Sin embargo existen 5 características que definen que deben tener los datos para poder ser útiles a las herramientas y técnicas asociadas con “*big data*”, es decir para ser considerados macrodatos, dichas características son llamadas las 5 v's del “*big data*”:

-Volumen: para que los datos generen posibilidades de modelización deben generarse en grandes cantidades, puesto que esto supone un crecimiento de la muestra a niveles cuasi poblacionales, lo que se traduce en un significativo aumento de la precisión de los modelos.

-Velocidad: esta característica es clave, puesto que en términos de computación moderna el análisis de datos es casi instantáneo a nivel individual, pero la información en su conjunto debe estar permanentemente disponible y actualizada para poder generar ventaja competitiva en la toma de decisiones.

-Variedad: se trata de una de las características en las que el “*big data*” basa su atractivo empresarial, la posibilidad de captar información de un mismo individuo de diversas fuentes, puesto que de esta forma se puede modelar al individuo desde múltiples perspectivas aportando valor en diversas áreas de la empresa. Generalmente esta variedad se clasifica en 3 grandes grupos en función de si tienen o no el mismo formato y en qué grado: estructurados, semiestructurados y no estructurados

-Veracidad: para un sistema de macrodatos el uso de datos fidedignos resulta imprescindible puesto que un volumen alto de datos incompletos, inexactos o directamente falsos, puede desviar la información, es por ello que resulta esencial establecer un filtrado de datos a través del uso de matrices de veracidad, sistemas de referencia cruzada que gracias a la variedad de fuentes permiten contrastar los datos obtenidos respecto a un individuo, generando datos de calidad.

-Valor: hace referencia a la utilidad que dichos datos reportan al modelo, es decir si nos encontramos en primer lugar ante un dato estadísticamente relevante para el estudio y en segundo lugar a la veracidad de dicho dato, estando esta segunda cuestión profundamente ligada a la variedad de estos, puesto que la multiplicidad de fuentes permite la construcción de matrices de características o datos que permiten contrastar su veracidad y por tanto interpretar su valor.

(Camargo Vega; Camargo Ortega y Joyanes Aguilar, 2014)

Por tanto y como definición final de “*big data*” o macrodato, podemos hablar de un conjunto de datos, cualitativos o cuantitativos, que expresados en gran volumen y velocidad, son contrastados a través de diversas fuentes dotándolos de valor conforme a su veracidad y relevancia estadística, los cuáles a través del uso de la computación moderna permite modelar y procesar la información e incluso su utilización autónoma por equipos informáticos.

## **CAPÍTULO 2**

### **OBJETIVOS DEL TRABAJO DE INVESTIGACIÓN**

#### **2.1 OBJETIVOS GENERALES**

A semejanza de lo brevemente explicado en el capítulo introductorio, el objetivo de este trabajo es un análisis de las posibilidades del uso de macrodatos y sus principales herramientas asociadas para la generación de ventaja competitiva en el entorno corporativo actual y en concreto en el entorno pyme nacional , partiendo para ello de conceptos básicos sobre *“big data”*.

#### **2.2 OBJETIVOS ESPECÍFICOS**

Puesto que la ventaja competitiva no engloba necesariamente la totalidad de la organización, sino que se produce en una o varias áreas concretas donde se genera un valor añadido respecto a la competencia, el estudio de la generación de dicha ventaja competitiva a través *“big data”*. Es por ello que para la realización de dicho análisis nos centraremos en distintas realidades de la empresa y las analizaremos de forma individual, a fin de obtener una visión más flexible e implementable fuera del marco académico de las posibilidades del análisis de macrodatos en las organizaciones actuales.

Para esto analizaremos los siguientes aspectos concretos de uso de *“big data”* en el entorno empresarial:

- Departamento de marketing: especialmente orientado al marketing digital
- Departamento de producción: mediante un enfoque en el análisis de demanda, gestión de recursos productivos y control de calidad.
- Departamento de recursos humanos: analizando la viabilidad técnica y legal del uso del análisis de macrodatos para la toma de decisiones de gestión de personal.
- En perspectiva global, la viabilidad de la implantación de sistemas de *“big data”* en pymes conforme a la ventaja competitiva generada.



## CAPÍTULO 3

### USO Y APLICACIÓN DE “BIG DATA” EN EL ENTORNO CORPORATIVO ACTUAL

#### 3.1 SECTORIZACIÓN

La utilización de herramientas de “*big data*” como hemos expuesto anteriormente requiere un inmenso volumen de datos, por tanto, es lógico pensar que este solo es utilizado por grandes corporaciones ya sea a nivel de facturación o de empleados. Sin embargo, no se puede apreciar una relación directa entre el tamaño de las organizaciones y la viabilidad de los sistemas de análisis de macrodatos, pero sí es un hecho cierto que existe una relación directa y exponencial entre el volumen de información generada y recogida por las organizaciones en sus interacciones con sus entornos y la implementación de estos sistemas.

Por tanto a la hora de analizar de forma sistematizada la adopción del análisis de macrodatos debemos huir de la clasificación tradicional de pequeñas, medianas y grandes empresas, si bien es cierto que el grado de recursos económicos disponibles es un factor importante de cara a la adopción de estos sistemas, debido a la minimización de riesgos que supone a la hora de afrontar el coste oportunidad y la inversión inicial, aunque no resulta un factor decisivo. Cuestión que podemos observar al analizar el “Dossier de indicadores sobre uso de “*big data*” por empresas de España y Europa” publicado en abril de 2020 por el Ministerio de Asuntos Económicos y Transformación Digital en colaboración con la Vicepresidencia Tercera del Gobierno de España , RED.ES y ONTSI. Como podemos ver en términos Europeos(UE28) no existe un ostensible diferencia entre pequeñas y medianas empresas en el porcentaje de utilización de estas herramientas , siendo en 2018 la diferencia de un 2% de usuarios entre las pequeñas empresas y las medianas empresas, sin embargo al introducir a las grandes empresas en la cuestión observamos una diferencia del 21% explicada por dos factores: en primer lugar la dilución de riesgos que su volumen les permite a la hora de implementar estas herramientas en conjunto con su mayor disponibilidad de recursos, pero sobre todo la mayor eficiencia de los sistemas de “*big data*” debido a que ese mayor volumen empresarial permite un rápido desarrollo o acceso al “*Know How*”<sup>1</sup> y un mayor volumen de captación de datos.

---

<sup>1</sup> Know how: activo intangible de la empresa que agrupa los conocimientos claves para el éxito de esta.

**Tabla 1.1. Indicadores sobre el uso de Big Data en Europa: empresas que analizan big data desde cualquier fuente de datos:**

**Indicadores sobre uso de Big Data en Europa**  
**Empresas que analizan big data desde cualquier fuente de datos**  
 Porcentaje de empresas  
 Datos para Alemania, Francia, Italia, Reino Unido, UE28 y España por tamaño de empresa

Tamaño de empresa	2016	2018
<b>Grandes empresas (250 personas empleadas o más), sin sector financiero</b>		
Francia	24	37
Alemania	17	34
UE 28	25	33
Italia	30	30
España	21	30
Reino Unido	35	
<b>Pequeñas empresas (10-49 personas empleadas), sin sector financiero</b>		
Francia	10	14
Alemania	5	13
UE 28	9	10
España	7	9
Italia	8	6
Reino Unido	13	
<b>PYME (10-249 personas empleadas), sin sector financiero</b>		
Francia	11	16
Alemania	5	14
UE 28	10	12
España	8	10
Italia	9	7
Reino Unido	15	
<b>Todas las empresas, sin sector financiero (10 personas empleadas o más)</b>		
Francia	11	16
Alemania	6	15
UE 28	10	12
España	8	11
Italia	9	7
Reino Unido	15	

Fuente: (Ministerio de Asuntos Económicos y Transformación Digital, RED.ES, and Observatorio Nacional de las Telecomunicaciones y de la SI 2020)

Sin embargo, en atención al dossier anteriormente mencionado existe un criterio absolutamente decisivo en el uso de herramientas de “big data”, el sector de actividad de las empresas. Si bien cabe resaltar que el sector financiero es el mayor usuario de estas herramientas a nivel global, su hegemonía en este campo es tan absoluta que queda excluida de estos indicadores, debido a su propia opacidad. No obstante podemos observar cómo el ranking de sectores de actividad lo lideran los sectores tecnológicos de la actividad económica, seguidos por el sector servicios y concretamente. En tercer lugar nos encontramos las actividades científicas y de investigación igualadas con las actividades logísticas. Podemos ver por tanto como aún las actividades industriales y de comercio tradicional presentan un bajo grado de adopción.

**Figura 1.2. Indicadores sobre uso de Big Data en Europa: empresas que analizan big data de cualquier fuente de datos por sector de actividad.**

<b>Indicadores sobre uso de Big Data en Europa</b>				
<b>Empresas que analizan big data de cualquier fuente de datos (excluyendo otras fuentes)</b>				
<b>Porcentaje de empresas</b>				
<b>Datos para UE 28 y España por sector de actividad</b>				
<b>Sector de actividad</b>	<b>2016</b>		<b>2018</b>	
	<b>UE 28</b>	<b>España</b>	<b>UE 28</b>	<b>España</b>
Reparación de computadoras y equipos de comunicación (10 personas empleadas o más)	10	7	12	27
Agencia de viajes; servicio de reserva de operador turístico y actividades relacionadas (10 personas empleadas o más)	19	13	20	26
Sector de las TIC (10 personas empleadas o más)	19	20	21	24
Información y comunicación (10 personas empleadas o más)	20	17	22	23
Transporte y almacenamiento (10 personas empleadas o más)	14	14	18	18
Electricidad, gas, vapor, aire acondicionado y suministro de agua (10 personas empleadas o más)	15	13	19	15
Alojamiento (10 personas empleadas o más)	14	17	13	13
Actividades profesionales, científicas y técnicas (10 personas empleadas o más)	9	9	12	12
Comercio minorista, excepto de vehículos automotores y motocicletas (10 personas empleadas o más)	8	7	10	10
Comercio mayorista y minorista; reparación de vehículos automotores y motocicletas (10 personas empleadas o más)	8	7	10	9
Actividades inmobiliarias (10 personas empleadas o más)	8	8	8	9
Comercio al por mayor, excepto de vehículos automotores y motocicletas (10 personas empleadas o más)	9	7	11	9
Construcción (10 personas empleadas o más)	7	6	10	8
Actividades de servicios administrativos y de apoyo (10 personas empleadas o más)	10	7	12	8
Fabricación de computadoras, productos eléctricos y ópticos, equipos eléctricos, maquinaria y equipo n.e.c, vehículos de motor, otros equipos de transporte ...	7	4	10	8
Fabricación (10 personas empleadas o más)	7	5	8	7
Actividades de alquiler y arrendamiento, empleo, seguridad e investigación, servicios a edificios y paisajismo, administración administrativa, soporte de oficina y ...	9	7	12	7
Fabricación de productos a base de: alimentos, bebidas, tabaco, textiles, cuero, madera, pulpa y papel; publicación e impresión (10 personas empleadas o ...	6	5	8	7
Fabricación de coque, petróleo refinado, productos químicos y farmacéuticos básicos, caucho y plásticos, otros productos minerales no metálicos (10 personas ...	8	7	10	7
Actividades de alojamiento y servicio de alimentos y bebidas (10 personas empleadas o más)	10	9	9	6
Fabricación de metales básicos y productos metálicos, excepto máquinas y equipos (10 personas empleadas o más)	6	4	6	4
Programación informática, consultoría y actividades relacionadas, actividades de servicio de información (10 personas empleadas o más)	20		23	
Fabricación de muebles y otras manufacturas; reparación e instalación de maquinaria y equipo (10 personas empleadas o más)	5		9	
Actividades de publicación; producción de películas, videos y programas de televisión, grabación de sonido y publicación de música; programación y difusión (10 ...	18		21	
Fabricación de equipos eléctricos, maquinaria y equipo n.e.c. (10 personas empleadas o más)	7		10	
Telecomunicaciones (10 personas empleadas o más)	21		24	
Fabricación de vehículos de motor, remolques y semirremolques, otros equipos de transporte (10 personas empleadas o más)	8		11	

Fuente: (Ministerio de Asuntos Económicos y Transformación Digital, RED.ES, and Observatorio Nacional de las Telecomunicaciones y de la SI 2020)

A tenor de estos datos podemos obtener dos conclusiones , el tamaño de la organización es importante de cara a la adopción de sistemas de análisis de macrodatos , sin embargo este no es un factor decisivo , como sí lo es su sector de actividad puesto que en gran parte depende de este el encaje entre las necesidades operativas de los propios sistemas de “*big data*” y la generación potencial de valor en el seno de la organización , siendo un rasgo común de los sectores que muestran mayor grado de adopción un alto grado de afinidad en cuanto a las características de los datos generados con las 5 V’s expuestas en el Capítulo I.

### **3.2 TIPOS DE DATOS UTILIZADOS PARA “BIG DATA”**

Partiendo de la tipología de datos expuesta anteriormente a modo de introducción: estructurados, semiestructurados y no estructurados, en atención al formato de los mismos. Existen 5 grandes tipos de datos en consonancia a la información que estos arrojan para los sistemas de “*big data*”:

- Los “*big transaction data*” o datos de grandes transacciones: datos de carácter semiestructurado o no estructurado y generalmente referentes a la operativa empresarial, obtenidos mediante sistemas *CRM*(*Customer Relationship Management*), *ERP*( *Enterprise Resource Planning*) o bien de inventarios o registros de ventas. Estos datos en lo que a nuestro enfoque empresarial se refiera contienen información respectiva al cliente.
- Redes sociales y páginas web: son datos de carácter generalmente personal y basados en el rastro que el usuario construye mediante el uso de internet.
- Biométricas: son datos obtenidos de forma estructurada y mediante mediciones especializadas de los rasgos físicos o conductuales de la persona.
- Datos de generación humana: si bien toda la información de los sistemas de macrodatos en algún momento de su vida tiene origen humano , este tipo de datos hace referencia a aquellos generados directamente por la acción del individuo, tales como una llamada , un correo electrónico o una compra; entendida como acción.
- Datos M2M o “*machine to machine*”: referentes a datos obtenidos por tecnologías que se conectan a terceros dispositivos, utilizándolos como sensores o medidores para la obtención de un gran volumen de información.

Si bien esta tipología creada por la consultora IBM se basa en el origen de los mismos.

### **3.3 METODOLOGÍA, TRATAMIENTO Y HERRAMIENTAS DE RECOLECCIÓN DE DATOS.**

Existen tres grandes metodologías para el tratamiento de datos para sistemas de “*big data*”. En primer lugar nos encontramos con la metodología KDD o “*Knowledge Discover Database*”, centrado en la extracción de información de un stock de datos , siendo de carácter iterativo e interactivo y que consta de cinco etapas:

En primer lugar la selección, fase en la que se seleccionan las variables de interés y si estas van a partir de muestras los datos que deben formar las mismas junto con su distribución estadística.

En segundo lugar, la fase de pre-procesamiento, durante la cual los datos se criban y adecuan para poder ser consumidos por la herramienta de análisis de datos.

Encontramos en tercer lugar la fase de transformación, durante la cual se utilizan métodos de reducción de la dimensión u otros métodos de transformación a fin de poder aplicar las técnicas de minería de datos seleccionadas.

En cuarto lugar encontramos la fase de minería de datos, donde se intentan resolver las incógnitas planteadas con las técnicas escogidas, se inicia la búsqueda de nuevos hallazgos y se intentan detectar patrones en los datos introducidos.

Y por último la fase de interpretación o evaluación en la que se extraen las conclusiones del estudio y posteriormente se evalúan los modelos de minería de datos utilizados.

La segunda gran metodología para el tratamiento de macrodatos es la SEMMA o “*Sample, Explore, Model and Assess*”; que está ligada al software SAS , específico para minería de datos y consiste de nuevo en 5 fases: en primer lugar la fase de muestreo, en esta etapa se obtienen muestras de los datos con carácter representativos para el análisis a realizar , pero de un tamaño adecuado para poder manipularlas en un tiempo y con un volumen de recursos razonables. Posteriormente pasamos a la fase de exploración en la que el software realiza un análisis preliminar de los datos para obtener unas primeras conclusiones sobre su morfología y tendencia, a fin de guiarnos sobre el camino a seguir. En tercer lugar, la fase de modificación, en la que se modifican los datos, se aplican transformaciones y se realizan selecciones para crear las variables ya orientadas al proceso de selección del modelo. Tras esto encontramos la fase de modelado, en la que aplicando los modelos disponibles de minería de datos se obtienen funciones de las variables elegidas como predictoras , lo que nos permite predecir el valor de la variable objetivo. Y por último la fase de evaluación, en la que se realiza un control de la utilidad y fiabilidad de las previsiones arrojadas por el modelo y se estima su rendimiento.

Como última metodología de tratamiento de datos encontramos el CRISP-DM o “*Cross Industry Standard Process for Data Mining*”, ligado al software SPSS Clementine que es un modelo desarrollado por IBM, si bien es compatible con cualquier otro software , basado en nodos y rutas. Esta última metodología es la más enfocada de las tres al mundo empresarial, debido a su flexibilidad y carácter cíclico.

Consiste en seis etapas, la primera y distintiva de este tipo de tratamiento de datos es la fase de entendimiento del negocio, cuyo objetivo es la definición del problema de minería de datos desde los objetivos y requerimientos del modelo de negocio. En una segunda etapa, llamada de entendimiento de datos estos se recolectan y pre-analizan a fin de detectar problemas de calidad de los mismos, así como se agrupan en subconjuntos preliminares y se obtienen las primeras conclusiones que servirán de guía para el resto del estudio. En tercer lugar se constituyen los ya definitivos conjuntos de datos a partir de los subconjuntos anteriores y los datos en crudo. A partir de estos conjuntos llamados “*datasets*” comenzará la cuarta fase del proceso, el modelado, consistente en la aplicación de las técnicas de modelado correspondientes y del calibrado de las mismas. En quinto lugar, durante la fase de evaluación se evalúan los modelos, así como su proceso de construcción en relación a los objetivos de negocio fijados en la etapa de entendimiento del negocio, para por último llegar a la etapa de despliegue en la que los modelos obtenidos se industrializan, generando un sistema o entorno productivo que se entrega al usuario para su implementación.

(Camargo Vega, Camargo Ortega, and Joyanes Aguilar 2014)  
 (“Procesos de Análisis de Datos”, n.d.)

Por tanto podemos extrapolar que independientemente de la metodología de tratamiento de datos utilizada, los sistemas de macrodatos pretenden obtener a partir de datos crudos conjuntos de datos válidos para la aplicación de herramientas de minería de datos para la construcción de modelos.

### **3.4 LIMITACIONES OPERATIVAS DE “BIG DATA”**

A la luz de las cuestiones anteriormente expuestas, las soluciones de “*big data*” presentan un potencial aparentemente ilimitado, sin embargo existen dos grandes barreras que lastran el ritmo de su expansión. En un primer grupo de limitaciones nos encontramos aquellas que podríamos definir como de carácter humano, principalmente la carencia de medios humanos cualificados que gestionen la complejidad de estos sistemas que encarece la mano de obra alrededor de estas herramientas lo que ligado a la falta de certeza respecto a la potencial ventaja competitiva generada y al desconocimiento de los propios sistemas hacen que desde la perspectiva empresarial las soluciones de análisis de macrodatos a menudo no presenten una relación riesgo-beneficio atractiva para el empresario. Por otro lado , su base en los datos generados por las personas en sus interacciones cotidianas traen a colación la cuestión de la privacidad , traducida en un complejo paradigma legal , que si bien será analizado en el capítulo siguiente , supone una barrera de carácter jurídico e importantes dimensiones al ritmo de crecimiento de estas herramientas en la gestión empresarial.

## CAPÍTULO 4 PARADIGMA LEGAL DEL ANÁLISIS DE MACRODATOS A NIVEL NACIONAL Y EUROPEO

### 4.1 REGULACIÓN DE BIG DATA EN EL ÁMBITO EUROPEO

La complejidad de los sistemas de macrodatos y su funcionamiento encuentra una equivalencia jurídica a la hora de su regulación, desde qué datos se pueden recolectar, a su tratamiento, análisis o en su último extremo qué tipo de decisiones se pueden o no tomar en el ámbito empresarial mediante sistemas de “*big data*”. Sin embargo es en la privacidad donde se centran los límites fijados por el legislador europeo, mediante el Reglamento de Protección de Datos en Europa (2016/679) o GDPR por sus siglas en inglés.

Este reglamento es pionero tanto por la materia regulada como por la incorporación por primera vez de un régimen sancionador que no solo tiene como objetivo a las empresas con sede europea, sino también a aquellas que operan desde fuera de la Unión Europea sobre ciudadanos europeos. Dicho régimen plantea dos modalidades de sanción, una primera de tramo fijo con multas que ascienden hasta 20 millones de euros o bien hasta un 4% de la facturación anual de la empresa sancionada, si bien también incorpora medidas correctivas menos lesivas para el interés empresarial como la advertencia o imposición del cese de almacenamiento de datos.

El citado Reglamento se erige sobre tres líneas maestras. En primer lugar, define las condiciones en las que una empresa podrá tratar datos personales del consumidor, para ello debe tener un fin legítimo y específico, recogiendo sólo los datos necesarios para ese fin y garantizando la seguridad de los mismos, así como su destrucción cuando dejen de ser necesarios. Su segundo pilar fundamental es la rendición de cuentas, articula el Sistema de Registro de Tratamiento de Datos, en el que las empresas de más de 250 trabajadores y que utilicen estas técnicas deben darse de alta y dejar constancia mediante inventario de sus datos disponibles, la seguridad que los protege y el tratamiento que reciben; si bien cabe como crítica a este punto del reglamento en primer lugar el uso del número de trabajadores como criterio que deja la puerta más que abierta a la elusión del registro por compañías que si bien realizan una abultada facturación en el sector no alcanzan dicho número y en segundo lugar la no obligatoriedad del mismo registro para pequeñas y medianas empresas desprotegiendo al usuario al considerar este modelo de responsabilidad simplemente recomendable. A su vez incluso las empresas u organismos públicos que realicen tratamiento de datos deberán nombrar un Delegado de Protección de Datos (DPD) que deberá tener formación específica en la materia y velará por el cumplimiento de la política de protección de datos, siendo a su vez el enlace con la autoridad competente en esta materia a nivel nacional. Y como último pilar del reglamento 2016/679 encontramos el consentimiento informado del usuario, deben saber quién, cómo y para qué se utilizan sus datos, requiriendo de un consentimiento explícito y revocable. Siendo esta revocación del consentimiento uno de los puntos más novedosos del Reglamento ya que a diferencia de la Directiva 95/46/CE a la que deroga, crea un cauce que permite al usuario primero solicitar copia de los datos generados, exigir su destrucción y pedir la baja del servicio en cuestión con la misma facilidad con la que se requirió el consentimiento inicial.

(Gómez,2016)

Podemos por tanto ver mediante un incremento de la protección del usuario convirtiendo su privacidad en el elemento central de la regulación europea, hasta el punto de incorporar por primera vez el ya conocido como “Derecho al olvido”. Por primera vez nos encontramos ante una regulación sistemática del rastro dejado por el usuario con el objetivo de reforzar la transparencia, el conocimiento y aceptación del usuario y el régimen de responsabilidad, pero también con una protección general de la persona física a su sometimiento a las decisiones algorítmicas surgidas de estos datos, conforme al Artículo 22 del citado reglamento, que aunque de redacción confusa permite escindir claramente que los datos personales y su tratamiento no pueden desembocar en decisiones con consecuencias jurídicas o de “significativa afectación” sin un análisis más profundo que el de la pura matemática.

Por tanto podemos concluir que el Reglamento 2016/679, es un giro humanizador dentro de la legislación del tratamiento de datos y por tanto de los propios sistemas *de “Big data”*, limitando su alcance a fin de preservar la privacidad y el estatus jurídico de la persona física, evitando que esta se vea sometida a un análisis absolutamente frío de sus datos personales.

#### **4.2 LEGISLACIÓN NACIONAL, DIRECTIVA 95/46/CE y REGLAMENTO 2016/679**

A nivel nacional encontramos la Ley Orgánica de Protección de Datos de Carácter Personal (LO 15/1999) que quedó derogada en 2018 por la Ley Orgánica de Protección de Datos Personales y Garantía de los Derechos Digitales, siendo esta una actualización de nuestra legislación interna al Reglamento 2016/679, cuyos fundamentos ya han sido tratados. No obstante podemos observar que si bien en clave nacional simplemente se actualiza la norma, ya dotada del carácter de Ley Orgánica y por ende protectora de un derecho fundamental recogido en el Título I de la Constitución de 1978, si bien de discutida alineación con los Artículos 15 o 18, cabe ser más razonable entender este carácter de derecho fundamental con el Derecho a la Intimidad garantizado por el Artículo 18.1.

La realidad de nuestra legislación nacional, no presenta mayor complejidad que la traslación del reglamento europeo 2016/719 y la creación de la Agencia Española de Protección de Datos (AEPD), si bien nuestra realidad jurídica fruto de una codificación incompleta, disemina el impacto de esta Ley Orgánica en otras múltiples disposiciones normativas, que han dado fruto a diversas controversias, siendo la más relevante el conflicto con la Ley Orgánica del Régimen General Electoral( LO 2/2011), al permitir por considerarlo de interés público a los partidos políticos la recopilación y creación de bases de datos de carácter ideológico, amparado por el Artículo 56 del Reglamento General de Protección de Datos, llegando dicho precepto a ser declarado inconstitucional por la Sentencia del Tribunal Constitucional 76/2019.

Sin embargo, al ser nuestra legislación en materia de tratamiento de datos una simple traslación de la regulación europea este tipo de fricciones son comunes y a menudo resueltas para su encaje por el propio legislador, el Tribunal Supremo o bien el Tribunal Constitucional. No existe por tanto en términos de contenido una producción legislativa interna en materia de protección de datos si no una primera traslación de la directiva 95/46/ce en forma de la LO 15/1999 y una posterior actualización de dicha ley para adecuarla al Reglamento Europeo 2016/679.

No obstante cabe preguntarse, si bien el legislador español mantiene la categoría jurídica de Ley Orgánica al actualizar la norma por qué el legislador europeo decide mutar la forma de directiva europea a reglamento. Esta cuestión, que podría ser objeto de análisis por sí misma, responde a dos grandes factores, para el legislador español la Ley Orgánica es un instrumento sumamente protector y que evita problemas competenciales y de colisión con los derechos forales en un estado plurilegislativo en múltiples materias; en cambio el legislador europeo se

ve forzado a seleccionar el reglamento como instrumento para la consecución de los objetivos de política legislativa, puesto que la directiva da un amplio margen de maniobra a los 27 estados miembros, puesto que les insta a implementar en su legislación interna normas para el cumplimiento de los objetivos marcados por la Directiva 96/45/CE , dando lugar a complejos escenarios en el desarrollo legislativo y sobre todo a una heterogénea legislación europea. Es por ello que con el Reglamento 679/2016, se pretende homogeneizar la normativa de todos los estados europeos, forzándolos con el reglamento como instrumento a adaptar sus normas internas al derecho europeo y no a producir su propio derecho al respecto, agilizando el proceso de regulación del tratamiento de datos y adaptándolo mediante una regulación más homogénea a la realidad global.

Si bien en clave nacional y debido a las realidades parlamentarias y a la mayor facilidad de esto, España optó en todo momento por ceñirse a las líneas marcadas por Bruselas en materia de protección de datos, debido a la concordancia con nuestros principios constitucionales , elección que posteriormente le ha ahorrado la tortuosa reforma de una ley orgánica en una realidad parlamentaria mucho más fragmentada. Pudiendo por tanto observar que aún carente de iniciativa en cuanto a contenido la regulación en materia de protección de datos nacional presenta una línea notablemente más coherente en términos de política legislativa que la propuesta por el legislador europeo.

#### **4.3 INCIDENCIA DEL PARADIGMA LEGAL EN EL POTENCIAL DE “BIG DATA”**

Cabe a raíz de lo expuesto, preguntarse si este paradigma legal de protección del usuario supone una limitación al uso o al potencial de las herramientas de “*big data*” en el entorno empresarial. La realidad es que sí, el Reglamento 2016/679 supone por primera vez una definición del entorno operativo de los sistemas de “*big data*” que hasta su entrada en vigor era casi inexistente , más allá del primer esbozo hecho por la directiva 95/46/CE cuyo efecto quedó claramente diluido por la necesidad de legislación nacional. Si bien estas limitaciones no afectan como tal a la potencial utilidad de los sistemas de macrodatos, sino que imponen a las empresas usuarias el deber de implementar sistemas paralelos de control de estos, que garanticen la seguridad de la persona física usuaria en materia de privacidad, de las consecuencias de su interacción con las plataformas de recolección de datos o sobretodo del ejercicio efectivo de su “derecho al olvido”, es decir de la destrucción de sus datos.

Por tanto, lejos de un límite operativo en su acepción más estricta, la actual legislación configura un entorno operativo acotado. No centrándose en frenar estos sistemas limitándolos en cuanto a ámbitos decisionales o categoría de información recabada, por ejemplo; sino centrándose en la incidencia que dichas cuestiones pueden tener para el usuario del servicio como persona física.



## CAPÍTULO 5 LA UTILIDAD EMPRESARIAL DE BIG DATA

### 5.1 LA ORGANIZACIÓN DIRIGIDA POR DATOS

En el cada vez más volátil y complejo entorno empresarial, surge “*the data driven organization*”; la organización basada en datos donde el conocimiento en sus múltiples formas se ha convertido en la base de la generación de utilidad, es aquí donde la enorme cantidad de información que el consumidor lanza en sus interacciones con las organizaciones juega un rol clave. El “*big data*” es una revolución en la gestión empresarial puesto que permite el análisis de inmensos volúmenes de información cuyo fin último es el respaldo y apoyo de las decisiones empresariales.

Sin embargo, esta cuestión no se antoja novedosa, el análisis de datos y la toma de decisiones empresariales en base a ellos no son disciplinas jóvenes en el mundo económico, puesto que tienen su origen en la comprensión de las necesidades del consumidor y de la propia organización. No obstante el análisis de datos tradicionalmente ha sido el puente hacia la comprensión e interpretación de estas necesidades de forma adecuada para generar una base de cara a la toma de decisiones de gestión; cabría entonces preguntarse: “¿Qué aportan los sistemas de análisis de macrodatos a la realidad empresarial?”- La toma de mejores decisiones, decisiones en primer lugar, respaldadas por un mayor número de observaciones y sometidas a un análisis objetivo, más rápidas, enfocadas al entorno al que se enfrentan y tomadas en base a un entendimiento global de las variables que afectan a la organización.

Es por tanto, que la novedad y generación de utilidad que los sistemas de macrodatos aportan a la toma de las decisiones empresariales provienen de una mejor justificación de estas permitido por el desarrollo tecnológico y de una mayor agilidad, concordantes con la tendencia de los entornos competitivos hacia una mayor complejidad y volatilidad, impulsados por el aumento de la globalización desde la década de 1990.

A pesar de esta mayor sincronía organizacional con su entorno competitivo existe una barrera básica, la necesidad de un cambio de cultura empresarial respecto a la toma de decisiones, en la que si bien no se excluye totalmente el factor y la supervisión humanas, ambos quedan relegados del rol decisonal como actores principales, siendo esta una nueva característica que surge en la organización basada en datos, en ella el papel de la persona es desplazado de la toma de decisiones puramente económicas y se centra en tareas como el liderazgo , la gestión del talento y la ejecución de la decisión basada en “*big data*”.

(McAfee y Brynjolfsson,2021)

## 5.2 ¿QUÉ CARACTERIZA A LA ORGANIZACIÓN “DATA-DRIVEN”?

El modelo de organización basada en datos puede sintetizarse en una frase: “No puedo gestionar lo que no puedo medir” Alphabet o Amazon son grandes exponentes de organizaciones basadas en datos. Las “*data driven organizations*” no solo recopilan y analizan datos en bruto, sino que los convierten en la columna vertebral de su modelo de negocio.

Existen cuatro características fundamentales de las organizaciones basadas en datos. En primer lugar, en su seno existe una democratización de la información, permitida por el “*big data*” puesto que gracias a este cualquier miembro de la organización puede obtener la información que requiere para sus tareas de forma manejable en umbrales de procesamiento humanos y a su vez puede incorporar información al sistema, generando un ecosistema de información en constante cambio y adaptación a las circunstancias, lo que aporta una enorme flexibilidad organizativa incluso a gigantes tecnológicos como los anteriormente citados. En segundo lugar, una cultura empresarial que considera la información como un activo, cuestión que requiere de líderes organizacionales comprometidos con esta idea y que a menudo, como podemos observar en Microsoft desde la llegada de Satya Nadella como CEO, implica un aplanamiento del organigrama a fin de un mayor contacto entre mandos y empleados a fin de transmitir el por qué de las decisiones. Como tercera característica encontramos la creación de una “literatura de datos”, mientras que en la organización tradicional las bases de datos juegan un rol de apoyo en las iniciativas de negocio en las “*data driven*” estas bases de datos y su análisis tienen un papel central, por lo que se debe capacitar al empleado para “*leer, escribir y comunicar datos en contexto, con una comprensión de las fuentes, los métodos y técnicas analíticas aplicadas, y la capacidad de describir la aplicación del caso y el valor comercial resultante*”(Rollings 2018). Por último, apreciamos una estructura interna altamente flexible, en la que los equipos de trabajo colaboran libremente y se realizan reuniones de análisis de datos entre distintos departamentos de la organización, a fin de obtener una visión de conjunto dentro y fuera del perímetro organizacional.

(Rollings,2018)

## 5.3¿QUÉ VENTAJA COMPETITIVA APORTA SER UNA ORGANIZACIÓN “DATA-DRIVEN”?

El enfoque de una organización basada en datos como hemos visto, no puede ser una cuestión departamentalmente aislada sino que debe contagiarse al conjunto de la organización, siendo esto un presupuesto básico de su efectividad.

Esta necesidad de un enfoque global encuentra justificación en que la principal ventaja de estas empresas radica en que su flexibilidad organizativa les permite capitalizar oportunidades y detectar tendencias de consumo a mucha más velocidad que la organización tradicional, debido al uso de modelos predictivos generados por el análisis de macrodatos; y por ende deben de ser capaces de adaptarse en su conjunto a estos cambios. Por otra parte los mismos modelos, basados en los mismos datos, no solo le permiten la explotación de oportunidades potenciales, sino que le generan un conocimiento más profundo de sus mercados permitiéndole una gestión perfectamente enfocada al cliente. Y por último supone un aumento de la eficiencia y la productividad, puesto que el uso de los datos generados internamente permiten la construcción de equipos de trabajos y una asignación de tareas acorde a las capacidades y recursos del capital de la organización.

Aunque para que esto sea posible, la transformación de una organización hacia una organización “*data driven*” debe surgir de una transformación de la cultura empresarial, encontramos tres líneas maestras en la ventaja de estas organizaciones sobre los enfoques tradicionales: una mayor sincronía con el entorno que les permite aprovechar oportunidades y detectar amenazas con mayor anticipación, un mejor enfoque al cliente y una mejor gestión de los recursos.

#### 5.4.¿CÓMO TRANSFORMARNOS EN UNA ORGANIZACIÓN “DATA-DRIVEN”?

En el mundo globalizado actual y en un momento de cambio propiciado por acontecimientos como el COVID19 o la Guerra de Ucrania, la volatilidad del entorno competitivo se ha hecho patente, si bien esto resulta incontrolable, la capacidad de adaptación de la organización es la única respuesta viable para su supervivencia. Si bien la organización basada en datos se ha revelado como el modelo más adecuado para ello, cabe preguntarse cómo puede cambiar un modelo empresarial tradicional hacia el modelo “*data driven*”.

Aunque cada día más empresas están adoptando soluciones de “*big data*” y en particular herramientas de “*business intelligence*”, esto no se traduce al trabajo de la mayoría de los empleados, mostrando ratios de penetración de estas herramientas a nivel global de entorno al 10% de las organizaciones a nivel global. Es por tanto que si bien, comenzamos a ver una adopción significativa de estas herramientas, no apreciamos una mutación de la cultura empresarial suficiente para maximizar sus beneficios.. Al fin y al cabo aunque las herramientas de “*business intelligence*” proporcionen respuestas y soluciones a los desafíos empresariales es el factor humano el que dirige el cambio. Es por ello que a pesar de la adopción de estas herramientas es necesario inculcar una cultura corporativa “*data driven*” para obtener la plena eficiencia de estos recursos.

Implantar esa cultura es el verdadero reto empresarial puesto que requiere no solo de una reorientación de los recursos organizacionales, sino de un cambio de mentalidad del conjunto de recursos humanos. No obstante, ¿Cómo podemos inducir ese cambio de cultura empresarial?

En primer lugar, es vital que los líderes de la organización apoyen el proyecto y se involucren de forma activa en él, creando en primer lugar un mando de cara a estas cuestiones, conocido en líneas generales como el CDO, “*Chief Data Officer*”<sup>2</sup>o director de datos. El papel de este puesto será en primer lugar coordinar transversalmente la implantación de herramientas y procesos en la organización a este respecto y por otra parte mantener comunicación continua con los “*stakeholders*”<sup>3</sup> de la compañía a fin de transmitir constantemente el mensaje de que el uso de sistemas de macrodatos no tiene como objetivo justificar las decisiones empresariales, sino ser su motor. Sin embargo resulta complejo acometer tales reformas sin oposición, de ahí la importancia de la capacitación del personal a través de sesiones con distinto grado de formalidad de talleres, cursos y otras herramientas formativas para evitar en el capital humano de la empresa la sensación de desplazamiento y a la par transmitirles la necesidad de su participación a fin de democratizar la información tanto vertical como horizontalmente en la compañía.

(Rollings 2018)

(Janoscchek,2021).

---

<sup>2</sup> Chief Data Officer: persona experta y capacitada en la gestión empresarial y de datos que ocupa un rol directivo con el objetivo de involucrar a toda la organización en la política de gestión de datos.

<sup>3</sup> Stakeholder: participante o persona interesada en la actividad de la organización y que por ende se ve afectado por la misma

Por otro lado, el proceso debe ser medido a fin de corregir desviaciones , para lo que debemos establecer indicadores de rendimiento o *KPI'S*<sup>4</sup> por sus siglas en inglés , pero a la hora de diseñar estos debemos tener en cuenta la importancia de la psicología colectiva de la organización y la tendencia humana hacia el camino de menor resistencia , por ello el propio sistema de indicadores debe permitirnos mostrar al menos en sus primeras etapas avances en la implantación que expuestos de forma pública facilitarán la alineación de intereses entre todos los miembros de la organización hacia la transformación en “*data driven*”, y permitiendo a su vez la aportación a esta mediante sugerencias , opiniones y comentarios. Por último, los aspirantes a organizaciones basadas en datos deben dar un giro en la transparencia, explicando hacia abajo en el organigrama las decisiones tomadas de forma analítica e involucrando a sus subordinados en las metas comunes.

Podemos por tanto observar, como el giro de una organización hacia convertirse en una organización basada en datos, tiene más que ver con la cultura empresarial y especialmente con la forma de gestionar el cambio en esta que con la incorporación técnica. Es por tanto un ejercicio de gestión del capital humano , haciendo comprender los aspectos positivos del cambio y sus ventajas tanto individuales como organizacionales y no una mera incorporación de herramientas técnicas, puesto que estas herramientas encuentran su utilidad limitada si sus usuarios no acometen las soluciones propuestas por el análisis de datos.

---

<sup>4</sup> Key performance indicator o KPI's: indicador del desempeño, es una medida del nivel de rendimiento de un proceso , habitualmente expresado en términos porcentuales conforme a un objetivo previamente fijado.

## CAPÍTULO 6

### LA OBTENCIÓN DE VENTAJA COMPETITIVA EN LA COMPAÑÍA BASADA EN DATOS

#### 6.1 GANANCIAS DE EFICIENCIA E INNOVACIÓN

A tenor de lo expuesto anteriormente cabe preguntarse si el esfuerzo requerido para transformar una compañía en “*data-driven*” se ve compensado en relación riesgo-beneficio. La cultura “*data driven*”, a través del procesamiento de datos nos empuja a una mayor disponibilidad de información y por tanto a una toma de decisiones más cualificada; esto se traduce de diversas formas en una mejora de la eficiencia de la compañía.

En primer lugar, supone un salto de calidad en la función de control dentro de la empresa puesto que la existencia de motivos objetivos para la toma de decisiones y de los modelos surgidos a través de la analítica “Big data” permite fácilmente observar desviaciones respecto a la planificación y a su vez localizar con facilidad las causas de estas. Tesis respaldada por Eryk Brynjolfsson, economista del MIT que postula que las empresas dirigidas en base a la filosofía “*data driven*” consiguen un aumento de la productividad de entre un 5 y un 6%.

Por otra parte, este mayor flujo de información, unida a la ya existente experiencia supone una potente combinación que a menudo torna en nuevos conocimientos, válidos para la implementación de procesos, ajustes para la optimización de la producción o una mejora de las interacciones con los “*stakeholders*” de la compañía.

Por tanto podemos esbozar que existen dos grandes líneas a través de las cuales la analítica mediante “*big data*” y sobretudo la filosofía de gestión “*data-driven*” generan una ventaja efectiva sobre los competidores: la eficiencia y la innovación. Sin embargo cabe preguntarse por campos concretos de aplicación empresarial.

Es en el área de marketing donde el análisis de datos, ha demostrado desde su mutación a “*big data*” mayor desarrollo de la mano de los gigantes tecnológicos norteamericanos permitiendo a través de las siguientes aplicaciones la generación de una verdadera ventaja:

- Motores de búsqueda y recomendación: nos encontramos ante una herramienta basada en el análisis del perfil de los usuarios y su comportamiento en las interacciones con las herramientas de recolección de datos, para esbozar que desea el cliente y de tal forma ofrecérselo de forma directa, evitando un esfuerzo de captación que no finalice en un beneficio para la empresa. Generalmente esta labor se lleva a cabo a través de una técnica conocida como “filtrado colaborativo” consistente en la confluencia de algoritmos de “*machine learning*” y “modelos de co-ocurrencia”
- Análisis de campañas de marketing: el mayor volumen de información manejado por la compañía sobre el consumidor permite a esta identificar y segmentar con mayor facilidad al cliente objetivo a través de patrones de comportamiento que mediante análisis clúster son segmentados. A su vez esta mejor identificación del cliente objetivo y sus patrones de conducta permiten medir objetivamente su sensibilidad a campañas de publicidad, fidelización o promoción de ventas.
- Análisis social: “*big data*” a través de sus herramientas de recolección de datos en redes sociales, permite hallar patrones de influencia desde las interacciones entre usuarios, sirviendo esto para verificar o incluso diseñar atributos atractivos para el producto, pero sobre todo para sus canales de ventas.

- Monitorización de redes: el internet que conocemos hoy día dado el insumo de datos que aporta es el campo óptimo de análisis mediante “*big data*”, puesto que el análisis de estos datos ya no solo es posible de forma interna, sino también desde el exterior de la red de la empresa, permitiendo identificar patrones de consumo, extrapolar el poder adquisitivo del usuario o incluso localizar.
- Análisis financiero: las técnicas de análisis de macrodatos sin duda han abierto una nueva era en el sector financiero, no solo por la capacidad para el procesamiento de increíbles volúmenes de datos sino por su capacidad para incorporar e incluso manipular a través de algoritmos de aprendizaje o clasificadores bayesianos el sentimiento del inversor. Los macrodatos han permitido por primera vez conocer la opinión pública respecto a un activo financiero de forma objetiva, modelar patrones de especulación e incorporar toda la información disponible al respecto para el análisis de riesgos financieros y la realización de análisis predictivos, hasta el punto del desarrollo de herramientas de “trading cuantitativo” capaces de especular en los mercados financieros de forma autónoma , generando un modelo de negocio como el de Renaissance Technologies, fondo de cobertura norteamericano pionero en estas cuestiones.
- Detección del fraude: el análisis de datos correlacionados y obtenidos de múltiples fuentes permiten su contraste sistemático y en tiempo real permitiendo la mejora de los mecanismos tradicionales de detección del fraude, dentro y fuera del mundo empresarial.
- Previsión del entorno: la posibilidad de contrastar múltiples datos heterogéneos, provenientes de múltiples fuentes y en tiempo real para generar modelos, nos permite la generación de previsiones del entorno, susceptibles a su vez a un análisis DAFO. El análisis de macrodatos nos permite identificar las amenazas y oportunidades que ofrece nuestro entorno y a su vez identificar las fortalezas y debilidades de nuestra organización respecto a este.

Si bien , estos campos de aplicación han sido seleccionados a modo de ejemplo, en un contexto empresarial sin actividad ni entornos definidos, encontramos que a día de hoy la aplicación de “*big data*” puede ser prácticamente universal: sanidad, agricultura, industria, lingüística... etcétera , sin embargo en todos los campos donde se aplican estas herramientas y especialmente en aquellos en los que se aplican bajo la filosofía “*data driven*”, la generación de ventaja competitiva proviene de dos grandes aspectos: eficiencia e innovación.

(Cano 2007)

(Mayer-Schönberger y Cukier 2016,)

(Rollings,2018)

(McAfee and Brynjolfsson 2021)

## 6.2 CONCEPTOS TÉCNICOS ALREDEDOR DE “BIG DATA”

Aunque la naturaleza de este trabajo es ciertamente económico-jurídica, es inevitable de cara a unas futuras conclusiones introducir ciertos conceptos técnicos de índole básica en torno a las herramientas de análisis de macrodatos.

Al tratar la analítica de macrodatos, tratamos con tres conceptos principales e íntimamente relacionados: “*data warehousing*”, “*no-sql*” y “*cloudcomputing*”.

En primer lugar, al referirnos al “*data warehousing*”, nos referimos al conjunto de arquitecturas que permiten la transformación de los datos brutos en información útil. Es decir, con este concepto que en su traducción literal significa almacenamiento de datos, nos referimos al conjunto de software y hardware que contienen los datos brutos que serán a lo largo de su estancia en esta infraestructura clasificados y transformados en información útil, compartimentada en “*data-marts*”. Siendo este el primer requisito técnico para la operación de un sistema de analítica de macrodatos. Radicando su principal diferencia con la base de datos tradicional en la estandarización automática de la información y su fin analítico, en lugar de la estandarización manual y su fin de simple consulta de los datos.

En segundo lugar, las herramientas “*NO-SQL*”, es decir bases de datos no relacionales, que permiten almacenar datos no estructurados y clasificarlos en función de cuatro modelos para su posterior clasificación y procesamiento a través de herramientas basadas en map reduce, cuestión que trataremos más adelante. Siendo estos cuatro modelos de datos NO-SQL:

- Key value: modelo basado en la asociación de los valores disponibles a llaves para su acceso directo, en función de una característica observada, siendo el modelo básico.
- Document-oriented: modelo basado en el propio modelo key value, pero asigna los accesos directos en función de la estructura de codificación de datos observada en los valores.
- Column-oriented o Table-based: modelo que permite almacenar los datos en tablas tridimensionales indexadas mediante la asignación de llaves a los valores, a la imagen del modelo Key value”
- Graph-oriented: modelo que permite almacenar y representar a las entidades en sí y a sus relaciones representando su abstracción matemática de forma gráfica.

Es una nota importante al concepto de NO-SQL entender que estas en su dinámica de trabajo se ven afectadas por el llamado teorema de CAP o Conjetura de Brewer, que sostiene que de las tres características en un sistema de cómputo distribuido: consistencia, disponibilidad y tolerancia a la partición; sólo es posible garantizar dos de forma simultánea. Puesto que la elección de una de estas, condiciona de forma eliminatoria la existencia de una de las dos restantes. Por ejemplo, al elegir la disponibilidad, no es posible mantener la tolerancia a la partición puesto que estos datos están en uso y la red de computación no puede modificarlos. Siendo esta cuestión relevante, puesto que la modificación de los sistemas NO-SQL en el nivel de avance técnico actual requiere de la renuncia a una de las tres características y por ende la renuncia para la modificación de esta herramienta a la operatividad del sistema analítico en ese momento.

Y por último, nos encontramos ante el “*cloud computing*”, una tecnología que permite el acceso remoto a software, almacenamiento de archivos y procesamiento de datos, permitiendo usar los equipos de otra entidad en lugar de los propios. Siendo esto relevante, porque en la actualidad el “*cloud computing*” permite la subcontratación de los sistemas de “*big data*” a un proveedor, liberando al usuario del sistema de la carga técnica que supone la implementación del sistema y haciéndolo más accesible y democrático.

(Tabares Pérez y Hernández,2014)

### **6.3 MAP REDUCE, HADOOP Y SPARK**

Anteriormente hemos tratado las bases de datos no relacionales (NO-SQL) y hemos mencionado su relación con Map Reduce. Map Reduce es un modelo de programación diseñado como soporte a las operaciones sobre grandes colecciones de datos , siendo desarrollado por la compañía Yahoo tiene como objetivo la resolución de algoritmos en conjuntos de datos de gran tamaño(petabytes) contenidos en sistemas de archivos distribuidos(HDFS), es decir el “*data warehouse*” anteriormente tratado.

Si bien Map Reduce, desde 2012 funciona en manos de Apache Software Foundation, que ofrece en forma de software libre y de código abierto, de forma que cualquiera puede acceder e incluso modificar el código para adaptarlo a sus requerimientos, Hadoop.Cuando hablamos de Hadoop , hablamos de la concreción técnica de “Mapreduce”. Hadoop es un entorno de trabajo que permite programar aplicaciones para la resolución de estos algoritmos, caracterizados por la enorme cantidad de datos que requieren ser procesados. Cabe resaltar sobre Hadoop , el hecho de que esté programado en JAVA, un lenguaje de programación caracterizado por su sencillez, hasta el punto que es enseñado en la educación secundaria; cuestión que se debe a que en último término la resolución de estos algoritmos apenas se aleja del Teorema de Bayes.

Desde 2014 Apache Software Foundation, además ofrece Spark , un entorno de programación fruto de las modificaciones sobre el código de Hadoop realizadas por la Universidad de Berkeley, que lo convierte en hasta 100 veces más rápido al trabajar sobre la memoria RAM del ordenador, siendo además compatible con Python, Scala y R además del lenguaje de programación JAVA . Aunque a cambio de estas ventajas requiere de un hardware notoriamente más costoso.

(White,2012)

(Ríos,2019)

### **6.4 LA ADOPCIÓN DE LA FILOSOFÍA “DATA-DRIVEN” EN PYMES**

Según datos del DIRCE con fecha de julio de 2021, el 97% del tejido empresarial nacional se encuentra compuesto por pequeñas y medianas empresas, siendo el sector servicios la actividad predominante con un 73% de las empresas. No obstante antes de profundizar, se ha considerado pequeña y mediana empresa a toda empresa que tiene menos de 250 trabajadores, un volumen de negocio inferior a los 50 millones de euros y un balance por debajo de los 43 millones de euros. Nos encontramos ante empresas, que más allá de las características objetivas citadas, presentan ciertos rasgos distintivos en su funcionamiento. En primer lugar a menudo se encuentran limitadas a la hora de realizar actividades intensivas en capital o mano de obra; también suelen ser dirigidas por los propios socios , que se caracterizan por ser un número reducido de personas, configurando una estructura jerárquica de pocos niveles, y por último sistemáticamente se enfrentan al reto del crecimiento o la supervivencia.

No obstante partiendo de estos rasgos esbozados, cabe entender a una PYME como una estructura empresarial ciertamente frágil y sin capacidad financiera suficiente para incorporar herramientas como el análisis mediante *"big data"*. Sin embargo es precisamente esta estructura poco desarrollada y extremadamente sensible a su entorno lo que la convierte en ideal para la aplicación de herramientas de análisis de macrodatos. En primer lugar, porque *"big data"* permite una mejor orientación tanto de la estrategia comercial como de la gestión del conjunto empresarial permitiendo enfocar de forma óptima unos recursos escasos. Y por otra parte porque en un organigrama sumamente compacto e incluso individual permite la objetivar las decisiones.

De igual forma la juventud que en muchos casos caracteriza a las pymes y su número reducido de trabajadores, las convierte en ideales para la implementación de la filosofía *"data-driven"* en lugar de únicamente herramientas analíticas. Esto es debido en primer lugar al ambiente de trabajo más cercano que facilita la democratización de la información corporativa, así mismo su juventud significa un menor arraigo de una cultura empresarial previa y por último la combinación de ambos factores las dota de una alta flexibilidad potencial que se traduce en capacidad de adaptación al cambio.

Cabe preguntarse, si una empresa de pequeño o mediano tamaño maneja un volumen de datos que encajen en las 5v's suficiente como para el funcionamiento de las herramientas analíticas. La respuesta es sencilla, sí. Cualquier negocio por pequeño que sea a día de hoy dispone de datos suficientes, siendo la cuestión de estos la fuente, si bien la presencia en redes sociales o la tenencia de una página web son herramientas ideales para la captación de datos; el análisis de la estructura de costes o del tránsito de clientes son fuentes igualmente válidas. En la actualidad el *"cloud computing"*, ha permitido que PYMES subcontraten los datos a gigantes tecnológicos como Google o Amazon, haciendo que la preocupación por la implementación de un sistema de análisis de datos sea enormemente sencilla, simplemente comparar proveedores, alejando la complejidad de estos sistemas y los elevados costes de hardware para acercar la solución final a cambio de cantidades monetarias que apenas llegan a los 9 dólares al mes.

(*"Virtualización de escritorios persistentes nativos en la nube – Precios de Amazon WorkSpaces – Amazon Web Services"*, n.d.)

Encontramos por tanto que es adecuada la implementación de la filosofía *"data driven"* y con ella de las herramientas de análisis de macrodatos a las pequeñas y medianas empresas, puesto que en la actualidad su precio accesible y el desentendimiento de la complejidad técnica gracias a la subcontratación de estas soluciones, permiten una relación coste-beneficio positiva al uso de estas herramientas y sobretodo de este modelo de gestión.



## **CAPÍTULO 7**

### **BIG DATA ÉTICA Y RESPONSABILIDAD SOCIAL CORPORATIVA**

#### **7.1 EL FUTURO PARA “BIG DATA”**

Partiendo de que “*big data*” aún se encuentra en un estado inicial de adopción y al liderazgo en su implantación del sector tecnológico, pero también de su cada vez mayor aproximación a las estructuras estatales, resulta evidente que crecimiento del uso de las técnicas de análisis de macrodatos solo puede seguir el signo ascendente que se viene observando desde el año 2016

(Ministerio de Asuntos Económicos y Transformación Digital, RED.ES, and Observatorio Nacional de las Telecomunicaciones y de la SI 2020)

#### **7.2 ÉTICA, RESPONSABILIDAD SOCIAL Y PROPUESTA REGULATORIA**

A tenor de la previsible expansión de las técnicas de análisis de macrodatos, cabe preguntarse por el impacto que la recopilación, análisis y utilización de estos datos tendrá en el futuro de nuestra sociedad, más allá del surgimiento de nuevas técnicas, profesiones o negocios , la preocupación debe centrarse en el impacto negativo que este rastro de información personal puede suponer al ser convertido en activo intangible.

Cuestiones como la privacidad, la propiedad intelectual o la libertad de expresión se encuentran en el ojo del huracán respecto a un conjunto de herramientas cuya capacidad de monitorización, análisis y predicción permite con facilidad un uso poco adecuado para los intereses del conjunto social. Sin embargo, el rápido desarrollo de la tecnología contrasta con el lento avance de la realidad legislativa; si bien la regulación a menudo va por detrás de la realidad esta diferencia aumenta su tamaño al enfrentarse a una realidad tan ágil como la tecnológica y a pesar de esto la regulación anteriormente analizada aún se ve más sobrepasada por el desconocimiento del usuario.

El regulador se encuentra ante una cuestión enormemente compleja, proteger al usuario sin socavar el potencial beneficio para la sociedad que suponen estas herramientas. A raíz de las páginas anteriores, cabría identificar al universo del “*big data*” con un máquina invencible que todo lo ve y todo lo sabe, no obstante ¿Cómo se puede proteger al usuario de un elemento de tales características, sin recurrir a la prohibición?

Bien, la respuesta es tan compleja como la tarea del legislador, la única forma de encarar esta cuestión por lo basto de la materia no es únicamente mediante una regulación técnica de la propia cuestión como se está desarrollando actualmente. Sino mediante una regulación exhaustiva de la Responsabilidad Social Corporativa, a fin de garantizar las sostenibilidad social y ambiental del uso de herramientas de “*big data*”.

Esta regulación debería ir suponer una translación de 2 grandes ejes. En primer lugar una mejora de la seguridad para el reforzamiento de la confianza. Por otra parte, mantener una gestión de la información sensible mediante el establecimiento de marcos internos adecuados en base a la generación de estándares de control.

Para la implementación de este primer objetivo, la propuesta más aceptada es la creación de un estándar europeo de *API's*<sup>5</sup>, al que las entidades públicas y privadas deban someterse garantizando que las interacciones de los usuarios con ellas a través de estas aplicaciones cumplen un estándar mínimo de seguridad; en segundo lugar facilitar la portabilidad de datos, homogeneizando el ecosistema de programas de tratamiento de información y el formato de los mismos, siendo esto una extensión del ya existente derecho acceso y rectificación recogido por la Ley Orgánica de Protección de Datos. Por último, la creación de e-estado, un estado conectado, a través de la implementación europea de un equivalente al “*blue bottom*” norteamericano, que permite al ciudadano acceder a los datos personales de los que el estado dispone sobre él.

Para la implementación de este segundo eje y dado el carácter volátil del desarrollo tecnológico es vital acortar la brecha entre regulación y realidad. Para ello es necesario la creación de un órgano a nivel europeo de carácter permanente con dos funciones, en primer lugar la monitorización de las actividades relacionadas con análisis de macrodatos, a fin de poder identificar tendencias e iniciativas necesitadas de regulación y en segundo lugar el contacto permanente con los *CDO's* y “*Compliance Officers*”<sup>6</sup> de las entidades que llevan a cabo estas actividades a fin de poder recoger propuestas normativas e intercambiar información sobre la actividad regulatoria.

Podemos por tanto concluir que si bien los retos a los que se enfrenta la sociedad moderna con una herramienta como el análisis de macrodatos de cara a la protección del individuo y sus libertades son enormemente complejos, existen dificultades para una regulación efectiva por la amplitud de la materia y su constante avance, cabiendo sólo una regulación marco que defina unos mínimos de acción, no solo salvaguardados por el principio punitivo sino también por una colaboración de las entidades públicas y privadas a través de políticas de responsabilidad social corporativa.

(Zubitur Sasturiain,2015)

---

<sup>5</sup> *API's*:interfaz de programación de aplicaciones basado en un conjunto de definiciones y protocolos usados para diseñar e integrar el software de aplicaciones, permitiendo que unos productos y servicios se comuniquen con otros sin necesidad de saber cómo está implementado.

<sup>6</sup> *Compliance officer*: persona que se asegura que una empresa cumpla con los requisitos legales externos, así como con sus políticas internas.

## CAPÍTULO 8

### CONCLUSIONES

A tenor de lo expuesto y a pesar de encontrarse aún en una fase embrionaria, la tecnología de análisis de macrodatos y en especial en su combinación con la filosofía *"data-driven"*, es una importante fuente generadora de ventaja competitiva. En primer lugar por las potenciales ganancias en eficiencia e innovación que este modelo de gestión aporta; pero también por sus costes de implementación relativamente bajos y que permite una mayor sincronía tanto a nivel estratégico como operativo a la organización, dotándola de unas capacidades de previsión ya adaptación sin precedentes; y por último por el nuevo nicho de mercado que la implementación de sistemas de análisis de macrodatos suponen por sí mismos.

Podemos por tanto concluir, que al hablar del análisis con *"big data"* no lo hacemos solo de un conjunto de herramientas técnicas, sino del precursor de un nuevo modelo de gestión empresarial de creciente adopción, a pesar de los retos tecnológicos pero sobretodo legislativos que esta nueva era de la gestión empresarial presenta. Analizando los factores diferenciadores de este modelo de gestión basado en datos, cabe destacar su capacidad transformadora del tejido competitivo empresarial y por ende del mismo tejido social. Por ello a pesar de que es clara la infinidad de ventajas que el uso de estas herramientas y sobre todo en combinación con la filosofía *"data driven"* aportan, es importante coordinar los frutos de este modelo de gestión con los objetivos estratégicos de la empresa a fin de lograr una ventaja competitiva efectiva. Queda por tanto patente que incluso en pequeñas y medianas empresas, más susceptibles a su entorno y con menor capacidad de inversión, el análisis con herramientas de *"big data"* sí genera una verdadera ventaja efectiva, permitiéndoles corregir sus debilidades. Es por ello que podemos afirmar con contundencia que tanto el uso de herramientas de análisis de macrodatos, como la filosofía de gestión empresarial basada en datos, constituyen fuentes generadoras de ventaja competitiva.



## CAPÍTULO 9 BIBLIOGRAFÍA

1. Burrell, Jenna (2016). "How the machine 'thinks': Understanding opacity in machine learning algorithms." *Sage Journals* 3, no. 1 (Enero): 12.
2. Cano, Josep Lluís (2007). "Business Intelligence: competir con información." Madrid, Madrid: Banesto, Fundación Cultur.  
[https://d1wgtxts1xzle7.cloudfront.net/61128354/Business\\_Intelligence\\_competir\\_con\\_informacion20191104-55292-drxoy-with-cover-page-v2.pdf?Expires=1646681827&Signature=PidoPnNHnwgO7~D3gR8ymqx5EAYbrWcxRjQTWDXs~87fXQS7WGesa0ybqL93LIZxQ6lsVLvaasFR3tQxYWoV8lme](https://d1wgtxts1xzle7.cloudfront.net/61128354/Business_Intelligence_competir_con_informacion20191104-55292-drxoy-with-cover-page-v2.pdf?Expires=1646681827&Signature=PidoPnNHnwgO7~D3gR8ymqx5EAYbrWcxRjQTWDXs~87fXQS7WGesa0ybqL93LIZxQ6lsVLvaasFR3tQxYWoV8lme).
3. Camargo Vega, Juan José, Jonathan F. Camargo Ortega y Luis Joyanes Aguilar (2014). "Conociendo Big Data." *Revista Universidad Pedagógica y Tecnológica de Colombia* vol.24, num.38.
4. Directiva 95/46/CE del Parlamento Europeo y del Consejo, de 24 de Octubre de 1995, relativa a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos. (D.O.C.E, núm.281, de 23 de noviembre de 1995, páginas 31 a 50.)
5. Gómez, Roberto M. (2016). "CONTENIDO Y NOVEDADES DEL REGLAMENTO GENERAL DE PROTECCIÓN DE DATOS DE LA UE (REGLAMENTO UE 2016/679, DE 27 DE ABRIL DE 2016)." *GABILEX* 6, 6 (Junio): 25.
6. Janoschek, Nikolai (2021). "Pervasive BI - Dream or Reality? The BI Survey data gives the answer." BI Survey. <https://bi-survey.com/bi-deployment>.
7. Ley Orgánica 15/1999, de 13 de diciembre, de Protección de Datos de Carácter Personal (LOPD) (B.O.E núm.298, de 14 de diciembre de 1999.)

8. Ley Orgánica 372018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales.(B.O.E núm.294, de 6 de diciembre de 2018, páginas 119788 a 119857).
  
9. Mayer-Schönberger, Viktor y Kenneth Cukier(2016) “*Big data: La revolución de los datos masivos.*” Traducido por Antonio J. Iriarte Jurado:Turner.  
<http://catedradatos.com.ar/media/3.-Big-data.-La-revolucion-de-los-datos-masivos-Noema-Spanish-Edition-Viktor-Mayer-Schonberger-Kenneth-Cukier.pdf>.
  
10. McAfee,Andrew, y Erik Brynjolfsson.(2021): “Big Data: The Management Revolution.” *Harvard Business Review* 10, no. 90 (Octubre): 9.
  
11. Ministerio de Asuntos Económicos y Transformación Digital, RED.ES, y Observatorio Nacional de las Telecomunicaciones y de la SI. 2020. “Dossier de indicadores sobre el uso de Big data por empresas de España y Europa.”  
<https://www.ontsi.es/sites/ontsi/files/2020-06/DossierIndicadoresBigDataAbril2020.pdf>.
  
12. “Procesos de Análisis de Datos.”Big Data International Campus. Consultado 19/03/22  
<https://www.campusbigdata.com/big-data-blog/item/104-procesos-de-analisis-de-datos>.
  
13. Reglamento Europeo 2016/719 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos y por el que se deroga la Directiva 95/46/CE(Reglamento general de protección de datos)(D.O.U.E núm.119, 4 de mayo de 2016, páginas 1 a 88.)

14. Ríos, Marcos M.(2019):“Spark vs Hadoop, ¿cuál es mejor? | ESIC.” ESIC BUSINESS & MARKETING SCHOOL.Consultado 05/04/22  
<https://www.esic.edu/rethink/tecnologia/spark-vs-hadoop-cual-es-mejor>.
15. Rollings, Mike(2018):“Build a Data Driven Organization.” Gartner.  
<https://www.gartner.com/smarterwithgartner/build-a-data-driven-organization>.
16. Tabares Pérez, Luis F. y Jhonathan F. Hernández. (2014):“Big Data Analytics: Oportunidades, Retos y Tendencias.” *Universidad San Buenaventura* 20, no. 1 (Diciembre): 21.  
[https://d1wgtxts1xzle7.cloudfront.net/38520697/Tabares\\_Hernandez\\_2014-big\\_data\\_analytics\\_FINAL-with-cover-page-v2.pdf?Expires=1650401008&Signature=Ris8mh82hVbSSG8mYwln8TF-C24grM8WXIHatFvniullHe0VOtJLaZzPgeSAuaDJRrBqD6Kd~JGO9ww1eHzEGuuvf38l7F~hU5PVVd3TNtZyF](https://d1wgtxts1xzle7.cloudfront.net/38520697/Tabares_Hernandez_2014-big_data_analytics_FINAL-with-cover-page-v2.pdf?Expires=1650401008&Signature=Ris8mh82hVbSSG8mYwln8TF-C24grM8WXIHatFvniullHe0VOtJLaZzPgeSAuaDJRrBqD6Kd~JGO9ww1eHzEGuuvf38l7F~hU5PVVd3TNtZyF).
17. “Virtualización de escritorios persistentes nativos en la nube – Precios de Amazon WorkSpaces – Amazon Web Services.” n.d. Amazon AWS.Consultado el 20/04/22  
<https://aws.amazon.com/es/workspaces/pricing/>.
18. White, Tom.(2012)*Hadoop: The Definitive Guide*. N.p.: O'Reilly.
19. Zubitur Sasturiain, Iñigo(2015):“Big Data e Internet de las Cosas en la Responsabilidad Social Empresarial.” BIG DATA e IOT: EL AVANCE TECNOLÓGICO EN LA RESPONSABILIDAD SOCIAL EMPRESARIAL.  
[https://www.researchgate.net/profile/Iñigo-Zubitur-Sasturain/publication/292656177\\_Big\\_Data\\_e\\_Internet\\_de\\_las\\_Cosas\\_en\\_la\\_Responsabilidad\\_Social\\_Empresarial/links/56b09a9e08ae9f0ff7b5dc3e/Big-Data-e-Internet-de-las-Cosas-en-la-Responsabilidad-Social-Empres](https://www.researchgate.net/profile/Iñigo-Zubitur-Sasturain/publication/292656177_Big_Data_e_Internet_de_las_Cosas_en_la_Responsabilidad_Social_Empresarial/links/56b09a9e08ae9f0ff7b5dc3e/Big-Data-e-Internet-de-las-Cosas-en-la-Responsabilidad-Social-Empres)

