



Facultad de Ciencias Económicas y Empresariales

Grado en Administración y Dirección de Empresas (E-2  
Bilingüe)

## Trabajo Fin de Grado **El valor estratégico del dato**

El impacto de la tecnología, el Big Data y su valor  
estratégico

Estudiante: Alicia Ríos Romero

Director: Lucía Barcos Redín

Madrid  
Junio 2023



*Declaro bajo mi responsabilidad, que el Proyecto presentado con el título EL VALOR ESTRATÉGICO DEL DATO en el Instituto Católico de Administración y Dirección de Empresas - ICADE de la Universidad Pontificia de Comillas en el curso académico 2022-2023 es un trabajo original y único. El proyecto no es plagado de otros y la información ha sido extraída de otros documentos debidamente citados.*

*Fdo.: Alicia Ríos Romero*

*Fecha: 09/06/2023*

## **Agradecimientos**

*En primer lugar, me gustaría expresar agradecimiento a mi tutora, Lucía Barcos Redín, por haberme tutorizado y ayudado durante este último año a llevar a cabo este Trabajo de Fin de Grado. Lucía me ha guiado en el desarrollo de ideas y estructura del trabajo final, aportando comentarios y opiniones que me han ayudado a establecer una dirección para la realización del trabajo. En segundo lugar, me gustaría agradecer a la Universidad Pontificia de Comillas por haberme facilitado los recursos necesarios para la ejecución del Trabajo de Fin de Grado. También me gustaría agradecer el apoyo de mi familia y amigos que han realizado un papel muy importante ayudándome a superar los obstáculos en el camino.*

## EL VALOR ESTRATÉGICO DEL DATO

**Autor: Alicia Ríos Romero**

Director: Lucía Barcos Redín

Entidad: ICADE – Universidad Pontificia Comillas

### Resumen del proyecto

**Palabras clave:** Big Data, Analytics, valor, estrategia empresarial, tecnología, Duolingo, reCAPTCHA, impacto, usuarios, consumidores, empresas, organizaciones, modelo, datos, *insights*, análisis, decisiones, objetivos, oportunidades, infraestructura, privacidad, ética.

**Resumen:** En estos momentos no cabe duda, de que el mundo de los datos y la información se están convirtiendo en uno de los temas más importantes e influyentes en todos los aspectos de la vida económica, política y social. Este trabajo, se enfoca en demostrar la relevancia de la utilización de los Datos en el desarrollo y éxito de las empresas.

Al inicio del trabajo, se demuestra la importancia que tiene el avance tecnológico, que se ha producido en los últimos decenios, para poder almacenar, clasificar y analizar muchos datos, y así poder generar información útil. Esta información puede ser usada por las empresas, y así adaptar el producto o servicio que ofrecen al cliente, a las necesidades y gustos de este. Y este tratamiento de datos es lo que se conoce como *Big Data*. Es indudable que el *Big Data* añade un valor muy importante a las empresas permitiéndoles tomar decisiones con mucha más información, no solo interna de la empresa, sino también del mercado y la competencia.

Y para entender mejor como utilizan las empresas esta disciplina del *Big Data*, se han elegido dos empresas, Duolingo y reCAPTCHA que ofrecen a los clientes una manera fácil de aprender idiomas, y estos usuarios a su vez aportan valor a la empresa, ayudándoles en el aprendizaje y traducción de nuevas palabras. Ha sido posible hacer esto gracias al enorme avance tecnológico y la aparición de una herramienta tan potente como es el *Big Data*.

Por tanto, podemos concluir diciendo que estamos asistiendo a un momento en el que es crítico para todas las empresas reconocer “el valor estratégico del dato”, e incorporarlo como parte esencial de la empresa.

## THE STRATEGIC VALUE OF DATA

**Autor: Alicia Ríos Romero**

Director: Lucía Barcos Redín

Entity: ICADE – Universidad Pontificia Comillas

### Project summary

**Keywords:** Big Data, Analytics, value, business strategy, technology, Duolingo, reCAPTCHA, impact, users, consumers, companies, organizations, model, data, insights, analysis, decisions, goals, opportunities, infrastructure, privacy, ethics.

**Abstract:** There is no doubt that the world of data and information is becoming one of the most important and influential topics in all aspects of economic, political, and social affairs. This paper focuses on demonstrating the relevance of the use of data in the development and success of companies.

Early in the paper, it demonstrates the importance of technological progress, which has occurred in recent decades, to be able to store, classify and analyze many data, and thus be able to generate useful information. This information can be used by companies, and thus adapt the product or service they offer to the customer, to the needs and tastes of the customer. This data processing is what is known as Big Data. There is no doubt that Big Data adds a very important value to companies, allowing them to make decisions with much more information, not only internal to the company, but also about the market and the competition.

And to better understand how companies use this Big Data discipline, two companies have been chosen, Duolingo and reCAPTCHA that offer customers an easy way to learn languages, and these users in turn bring value to the company, helping them in the learning and translation of new words. It has been possible to do this thanks to the enormous technological advancement and the emergence of such a powerful tool as Big Data.

Therefore, we can conclude by saying that we are witnessing a moment in which it is critical for all companies to recognize "the strategic value of data" and incorporate it as an essential part of the company.



## Índice

Agradecimientos .....	iii
Resumen del proyecto .....	iv
Project summary .....	vi
Índice.....	viii
Índice de abreviaturas .....	x
Índice de tablas .....	xi
1. Introducción .....	1
1.1. Motivación .....	1
1.2. Objetivos .....	1
1.3. Metodología .....	2
1.4. Estructura .....	3
2. La tecnología.....	3
2.1. El impacto de la tecnología en las estrategias empresariales.....	3
2.2. Los beneficios tecnológicos .....	8
3. El Big Data.....	10
3.1. ¿Qué es el dato? .....	10
3.2. Tipos de datos.....	13
3.3. El dato y su ciclo.....	16
3.4. Marco de clasificación del Big Data .....	20
3.4.1. Elementos esenciales en Big Data para la creación de valor .....	20
4. El valor estratégico del dato .....	21
4.1. El concepto del valor.....	21
4.2. El beneficio económico del valor .....	23

4.3.	Captura de valor y creación de valor .....	24
4.4.	Marco de clasificación del valor .....	25
4.4.1.	El valor según su dirección .....	25
4.5.	Limitaciones en la creación de valor .....	29
4.5.1.	El coste de la infraestructura .....	29
4.5.2.	Los datos no estructurados .....	30
4.5.3.	Protección y privacidad de los datos .....	31
4.5.4.	Escasez de competencias y talento .....	33
5.	Caso real: Duolingo y reCAPTCHA .....	34
5.1.	Introducción .....	34
5.2.	Análisis empresarial .....	36
5.3.	Conclusiones del caso y recomendaciones .....	40
6.	Conclusiones .....	41
6.1.	Conclusión de objetivos .....	41
6.2.	Limitaciones .....	42
6.3.	Futuras líneas de investigación .....	42
7.	Referencias .....	44

## Índice de abreviaturas

<b>BDA</b>	<i>Big Data and Analytics</i>
<b>CAPTCHA</b>	<i>Completely Automated Public Turing test to tell Computers and Humans Apart</i>
<b>IoT</b>	<i>Internet of Things</i>
<b>SQL</b>	<i>Structured Query Language</i>

## Índice de tablas

<i>Tabla 1: Categorías de datos en modelo predictivo en el sector hospitalario</i> .....	15
<i>Tabla 2: Direcciones de valor de Duolingo y reCaptcha</i> .....	41

## 1. Introducción

A lo largo de la historia de la humanidad, la evolución tecnológica ha sido un proceso exponencial. Esta evolución ha desempeñado un papel vital en el desarrollo humano, generando cambios sociales y económicos, especialmente durante las revoluciones industriales. Las dos revoluciones más recientes (últimos 50 años), han producido un avance superior, a los avances previos desde el inicio de la humanidad: desde la automatización de procesos, la aparición de la telefonía móvil y digitalización, hasta la Inteligencia Artificial, Internet de las Cosas, etc. Esta evolución no solo es relevante para la vida de las personas, sino también para el mundo empresarial, que es el punto de enfoque de este trabajo.

### 1.1. Motivación

La motivación para elegir este tema ha sido porque en los últimos años los conceptos de *Big Data* y *Analytics* han ido creciendo en importancia, hasta convertirse en un tema de actualidad, presente en los medios y redes sociales. No todo el mundo entiende que hay detrás de ello, ya que es complejo para las personas que no tienen conocimiento técnico. Por ello, he querido profundizar en este tema, centrándolo en una parte que me interesa mucho como es la contestación a estas dos preguntas:

- 1- ¿Qué influencia está teniendo el *Big Data* en la empresa y que valor le aporta?
- 2- ¿Cómo puede ayudar el *Big Data* a la empresa en este entorno global cada vez más competitivo?

### 1.2. Objetivos

Por la velocidad que lleva la evolución tecnológica en los últimos años, parece que el futuro próximo va a traer muchos más avances, que todavía no somos capaces de imaginar. Para entender esta evolución, hay que tener en cuenta 3 elementos relacionados entre sí: la Tecnología, el *Big Data* y el Valor.

La tecnología en la empresa es un medio que permite mejorar los procesos de esta y hacer que la forma de trabajar sea cada vez más eficiente. De forma que, antes se necesitaban muchos más trabajadores que ahora, que todo está automatizado. Y esta automatización ha provocado que se tengan más datos e información.

Esta avalancha de información ha llevado a la creación del concepto de *Big Data*, ya que la tecnología ha sido un habilitador clave para la recopilación y procesamiento de Datos. El *Big Data* es una herramienta de análisis de datos, que posibilita el almacenamiento y procesamiento de grandes cantidades de Datos, permitiendo a las empresas tener una perspectiva más amplia acerca los clientes, su negocio y el entorno competitivo. Y se ha convertido en un medio para generar valor y crecimiento en las empresas.

Por tanto, el valor que aporta el *Big Data* en el ámbito empresarial es indiscutible, ya que las compañías pueden obtener información crítica en tiempo real de los consumidores, mercado y competidores, que les permite tomar decisiones estratégicas basadas en los Datos.

A través de los casos reales de Duolingo y reCAPTCHA, que se mencionan al final de trabajo, se analizará el valor extraído en dichas empresas a partir del uso del *Big Data*, creado por el reciente avance tecnológico.

### 1.3. Metodología

Para la realización de estudio se ha llevado a cabo una revisión de literatura, seleccionando artículos académicos de las siguientes fuentes:

- *Scopus*
- *Web of Science*
- *Google Scholar*

En estas fuentes se ha realizado una búsqueda inicial de artículos relacionados con el valor estratégico, el impacto tecnológico, el *Big Data* y la creación de valor en las empresas, entre otros.

De los artículos leídos se han seleccionado aquellos que se adaptan al objetivo de este trabajo, que es el demostrar “*el valor estratégico del dato*”.

La elección de los casos prácticos está inspirada en las entrevistas realizadas por el creador de ambas empresas, Luis von Ahn. Con estos casos se pretende ilustrar mediante empresas reales la posibilidad de aplicar el marco de trabajo de valor, basado en el conjunto de direcciones de creación de valor, que se enumeran y explican en el capítulo cuatro del presente documento.

#### 1.4. Estructura

Este trabajo este compuesto por siete capítulos. Tras la introducción, se describen los tres elementos que forman parte del objetivo del trabajo: la tecnología, el *Big Data* y el valor estratégico del dato. A continuación, se analizan los casos de Duolingo y reCAPTCHA. Finalmente se muestra la conclusión del trabajo y las referencias.

## 2. La tecnología

### 2.1. El impacto de la tecnología en las estrategias empresariales

Según la definición de la Real Academia Española (2023), se entiende por tecnología al conjunto de teorías y de técnicas que permiten aplicar los conocimientos científicos a la práctica. Estas teorías y técnicas son empleadas en todos los aspectos de la vida, pero especialmente tiene influencia en el área empresarial, donde condiciona el funcionamiento, desarrollo y viabilidad de las empresas. Además, la tecnología no es algo estático que permanece inalterable en el tiempo, sino que cambia y evoluciona constantemente, lo que afecta de forma directa e indirecta a la actividad empresarial.

Antes de profundizar en el impacto que tiene la tecnología en el mundo empresarial, es necesario determinar qué se entiende por producto digital, que es la base de la innovación tecnológica.

Según Nam y Lee (2003), los productos digitales se componen de elementos hardware (componentes físicos del producto, por ejemplo, chips, memorias, etc) y de elementos software (programas que implementan los algoritmos que se ejecutan sobre el hardware). Y es precisamente en estos dos elementos donde se están produciendo grandes avances e innovaciones, que están haciendo progresar las actividades económicas.

Para poder entender la innovación que están experimentando las actividades económicas en la sociedad, hay que comprender qué procesos se establecen para el desarrollo de nuevos productos o mejora de los ya existentes, tanto a nivel del hardware como del software de un producto. Estos procesos están basados en arquitectura modular. La arquitectura modular hace referencia al proceso por el cual, para desarrollar un producto, se hace una descomposición de este en diferentes estratos. Cada estrato ha de proporcionar una función concreta y cada una de esas piezas se van uniendo para formar el producto final, de acuerdo con el diseño definido por los arquitectos (Yoo, Henfridsson, y Lyytinen, 2010).

A continuación, se enumeran algunos ejemplos de cómo se han mejorado los productos a través de la innovación tecnológica:

- Las cámaras de los teléfonos móviles han incorporado elementos innovativos a la tecnología de las cámaras ya existentes, como módulos de procesamiento de imagen basados en Inteligencia Artificial, que permiten mejorar la calidad de la imagen a nivel profesional.
- Los *robots* de aspiración en el sector de los electrodomésticos han incorporado módulos tecnológicos que les permite hacer una aspiración autónoma y más eficiente, siendo capaces de memorizar los espacios a aspirar de la vivienda.
- En el sector financiero el análisis del riesgo crediticio en créditos al consumo se hacía prácticamente de forma manual, ahora se ha incorporado un módulo tecnológico basado en algoritmos de inteligencia artificial obtenidos del análisis de los datos agregados de muchos usuarios (Munafa, 2019).



Podemos decir que esta nueva forma de producción de bienes y servicios ha acelerado las innovaciones y estamos en estos momentos en una era digital, donde la tecnología está presente en todos los aspectos de nuestras vidas, ya sea a nivel personal o profesional. Está cambiando:

- La manera de trabajar con grupos formados por trabajadores de varios países que trabajan de forma virtual como un único grupo local, aprovechando las facilidades del teletrabajo
- La forma de socializar donde las redes sociales han adquirido mucha relevancia y ya nadie concibe la vida sin ellas
- La formación cada vez es más online, pudiendo acceder a información de la otra parte del mundo, etc.

Por tanto, la integración tecnológica en la arquitectura modular de los productos evidencia la importancia de que las empresas han de tener en cuenta la tecnología a la hora de formular estrategias empresariales de cara al futuro. Esto implica que las organizaciones deben reconocer el papel fundamental que desempeña la tecnología en su éxito y supervivencia a largo plazo. Una razón clave por la cual es importante que las empresas consideren la tecnología en sus estrategias empresariales, es la competencia, ya que las empresas que no utilizan tecnologías en desarrollo corren el riesgo de perder una ventaja competitiva sobre sus rivales a medida que el entorno empresarial se va digitalizando. Otra de las razones por las que la tecnología debe formar parte de la estratégica de una empresa, es porque la eficiencia operativa y la productividad de una empresa aumentan notablemente con la tecnología. El término de eficiencia y productividad en una empresa van asociados a la reducción de costes, que es clave en la cuenta de resultados de las empresas (Yoo, Henfridsson, y Lyytinen, 2010).

Por lo tanto, las futuras evoluciones y obstáculos que puedan tener las empresas podrán ser afrontados mejor por aquellas que estén preparadas para adaptarse y utilizar las nuevas tecnologías.

Históricamente, los productos se agrupaban bajo una industria determinada. Las empresas que lograban diferenciación en sus productos podían amortizar la inversión, ya que la competencia tardaba un tiempo en aparecer. Sin embargo, actualmente, nos encontramos en un entorno digital dinámico que evoluciona continuamente, y la distinción entre un producto y una industria se ha

vuelto más difusa. Los productos digitales no son estáticos, ya que deben ajustarse constantemente al entorno interno y externo de la empresa. (Yoo, Henfridsson, y Lyytinen, 2010).

Según Yoo Henfridsson y Lyytinen (2010), algunos conceptos como el ciclo de vida del producto no pueden utilizarse en un entorno digital porque se desarrollaron para productos fijos de carácter tradicional, que no se veían afectados por la tecnología digital. Además, se requiere una visión dinámica y fluida de los productos, que tenga en cuenta su continua evolución, actualización y personalización, en respuesta a las aportaciones de los clientes y las tendencias del mercado. Ahora esa diferenciación cuesta más mantenerla y los entornos se vuelven mucho más competitivos, con la necesidad de innovación continua.

Para desarrollar estrategias competitivas y nuevos marcos estratégicos para industrias y bienes cada vez más adaptables, hay que buscar nuevas propuestas de valor, que permitan diferenciarse del resto, pero esto conlleva inversión que no todas las empresas pueden permitirse, por tanto, es importante tener una capacidad financiera. Es necesario saber con qué recursos cuenta la empresa para llevar a cabo esa digitalización e innovación tecnológica. Además, esta estrategia digital tiene que venir acompañada de un plan estratégico de viabilidad para definir y analizar posibles limitaciones que se presenten en un futuro relacionadas con la implementación digital llevada a cabo (Elia et al., 2020).

Y lo que está ocurriendo con estos cambios tecnológicos, es que la velocidad de estos ha aumentado, especialmente en las últimas décadas, lo que lleva a que las empresas han de adaptarse a estas tecnologías emergentes, para alcanzar las expectativas de los clientes y poder seguir funcionando en un mundo que cambia cada vez más deprisa.

Por tanto, las empresas se pueden ver impactadas por dos tipos de factores:

- Los factores externos de cambios tecnológicos, que conllevan un impacto en la actividad empresarial y el funcionamiento de las compañías, al añadir nueva competencia en el mercado.
- Y los factores internos que pueden tener un efecto directo en las empresas, al ser las propias empresas las que adoptan una nueva innovación. Es el caso de las innovaciones

implementadas en la cadena de suministros de la compañía, que puede modificar la trazabilidad y distribución de los productos y servicios.

Esto demuestra el enorme impacto que tienen las tecnologías en la actividad empresarial. En el caso concreto del *Big Data y Analytics*, según Müller, Fay y Brocke (2018) se estima que las empresas que utilizan herramientas de BDA, han experimentado un incremento de la productividad de alrededor del 4,1 %.

Esto se traduce, en que la mayoría de las empresas cada vez son más conscientes de la importancia de las nuevas tecnologías en la sociedad y por lo tanto en su empresa. Sin embargo, no todas las empresas son conscientes de que las nuevas tecnologías se pueden emplear como activos para extraer valor y obtener información muy útil, a partir de los datos e información que hay en la empresa, como es el caso del *Big Data*. Esta información extraída, ordenada y analizada puede ser utilizada en la toma de decisiones por parte de la dirección de la empresa durante la elaboración de la estrategia empresarial y así conseguir una mejora significativa en la operativa de las empresas, y consecuentemente en los resultados financieros de la misma, que es el fin último de toda empresa.

Y lo que es aún más importante, que muchas de estas decisiones tomadas a raíz de los datos extraídos de la empresa y debidamente modelizados, pueden tener lugar en tiempo real, por lo que permitirá adaptar esas decisiones a las circunstancias existentes en ese instante concreto.

Mediante las soluciones BDA, las empresas pueden realizar análisis de grandes volúmenes de datos con el objetivo de encontrar posibles patrones, relaciones e incluso tendencias en la actividad empresarial, lo que puede repercutir en un mejor funcionamiento de sus procesos y una mayor productividad de sus empleados. Otra forma mediante la cual las soluciones BDA pueden mejorar la productividad empresarial es mediante la optimización de procesos internos y de sus operaciones en las distintas áreas empresariales.

Esta tecnología BDA también se puede aplicar a los clientes de la empresa, y es ahí donde se abre un gran abanico de posibilidades. Una de esas posibilidades, es que a través de la información que

haya de los clientes se podrán establecer patrones de comportamiento, y por tanto segmentar el mercado de acuerdo con ese patrón. Además, los productos o servicios que se suministren por parte de la empresa a esos clientes se podrán personalizar, lo que sin duda creará una mayor satisfacción por parte de esos clientes, y la fidelización de los mismos.

Y para analizar y poder usar esos datos que se están obteniendo han aparecido nuevas disciplinas tecnológicas en el mercado, que de momento no son muy conocidas para la mayoría de la población, pero que en un futuro próximo formarán parte de nuestras vidas (Gill et al., 2019):

- El Internet de las cosas (IoT)
- La computación en la nube
- El *blockchain*
- La inteligencia artificial (IA), el aprendizaje automático
- etc.

Todas estas nuevas tecnologías, permiten la puesta en acción de una estrategia empresarial enfocada en la mejora de productividad de la empresa, la mejor adaptación al cliente, la fidelización de este y lo que es más importante, la mejora en la cuenta de resultados de la compañía.

## 2.2. Los beneficios tecnológicos

La tecnología se ha convertido en un recurso imprescindible para las empresas a la hora de formular estrategias que sustenten la actividad empresarial y puedan llevar a cabo transformaciones en el proceso de toma de decisiones de la organización.

Si las inversiones tecnológicas que se llevan a cabo son las adecuadas para el tipo de organización, puede ofrecer gran cantidad de beneficios al rendimiento de la empresa.

Según Raguseo (2018), estos beneficios se pueden clasificar en cuatro grupos:

- **Beneficios estratégicos.** Los beneficios estratégicos hacen referencia a aquellas ganancias que pueden obtener las empresas con respecto a otros competidores y con respecto a las

mejoras en sus productos o servicios. Entre estos beneficios también destaca la innovación aplicada a procesos de los bienes y servicios creando oportunidades de negocio.

Por ejemplo, *Netflix* ha conseguido un posicionamiento relevante en la venta de contenidos, gracias al modelo innovador de recomendación de contenidos personalizados al usuario, basado en BDA e inteligencia artificial.

- **Beneficios informativos.** Los beneficios informativos son aquellos que proporcionan información de valor para la empresa y para mejorar el proceso de toma de decisiones. Cabe destacar como beneficio informativo el seguimiento del progreso mediante el cual se establece si la empresa está avanzando correctamente con respecto a los objetivos establecidos. Pero principalmente, los beneficios informativos van a ir encaminados a llevar a cabo decisiones sustentadas por conocimientos.

Por ejemplo, la compañía *UPS* dispone de mapas de datos geoespaciales y algoritmos predictivos en el reparto de mercancías, que permiten una optimización operativa y una mayor visibilidad en sus procesos logísticos (Grover et al., 2018).

- **Beneficios transformacionales.** Los beneficios transformacionales van a consistir en aquellas modificaciones que se van a realizar en las organizaciones y en su estructura. Más concretamente, se tratan de aquellas transformaciones digitales en las que se aplican las nuevas inversiones en tecnologías para mejorar los procesos en el negocio.

Es el caso de la compañía financiera *Deutsche Bank*, que ha desarrollado capacidades automáticas de detección de anomalías de transacciones financieras, detección de fraude y seguridad. Estas capacidades han permitido a la entidad conseguir optimizaciones operativas, apoyando la participación humana en este tipo de tareas (Grover et al., 2018).

- **Beneficios transaccionales.** Los beneficios transaccionales son aquellos que están relacionados con las inversiones llevadas a cabo por la empresa con el objetivo de mejorar la eficiencia y efectividad de los procesos operativos. En esta era digital va a destacar como

beneficio transaccional la automatización de procesos que permitirá a largo plazo una reducción de costes al igual que la agilización de transacciones empresariales.

Por ejemplo, la empresa *Inditex* ha mejorado su eficiencia operativa utilizando soluciones IoT y *Big Data*, debido a la situación provocada por el COVID-19 (González-Padilla y Galán Zori, 2021).

El elemento común en estas nuevas disciplinas, que las hace posible es la recopilación y análisis de datos (*Big Data*).

### 3. El Big Data

#### 3.1. ¿Qué es el dato?

Los datos son el conjunto de elementos obtenidos de una determinada materia, que debidamente agregados se obtiene información sobre este punto concreto.

Según Urbinati et al. (2019), *Big Data* se podría definir, por tanto, como la agregación de un conjunto de datos, que debido a su tamaño y complejidad no pueden ser analizados con técnicas convencionales de procesamiento de datos y que por lo tanto necesita la aplicación de algún algoritmo para la detección de elementos comunes entre estos datos. A esta aplicación de algoritmos para la obtención de patrones de comportamiento, consumo, etc, se le denomina *Analytics*. El conjunto de ambos términos, *Big Data* y *Analytics*, se le conoce como BDA (Grover et al., 2018).

Actualmente nos encontramos en una era digital en la que se ha producido un aumento exponencial de los datos a través de los medios tecnológicos, dando lugar al “*Big Data*”. Y esos datos se han convertido en un elemento muy valioso para todas las facetas de la vida humana.

Por ello, el surgimiento del *Big Data* o dato masivo ha llevado a las empresas a adoptar herramientas de análisis de negocios para poder llevar a cabo un análisis cuantitativo extensivo de

esos datos, creando modelos predictivos para poder facilitar la toma de decisiones de la empresa (Müller, Fay, y Brocke, 2018).

Las personas también nos estamos acostumbrando a usar herramientas digitales que requieren de BDA en nuestra vida cotidiana. Por ejemplo:

- Buscador de Google, que tiene con una simple palabra es capaz de encontrar resultados coherentes, relacionados con esa palabra, gracias a la indexación masiva de páginas Web por todo el mundo.
- Modelos de recomendación de plataformas como Amazon, usan *Big Data* prediciendo los productos a ofrecer a un cliente concreto.
- Las plataformas de video como Netflix o HBO, también agregan información para segmentar los clientes, etc.

Centrándonos en la importancia del dato, el *Big Data* se entiende mejor mediante las cinco dimensiones o modelos de “*Las 5 V’s del Big Data*” (Elia et al., 2020). Este modelo plasma cinco características creando un marco de referencia que ayudan a comprender las oportunidades y problemas que plantea el *Big Data*. Inicialmente se trataba de un modelo que contaba con tan solo tres dimensiones: volumen, variedad y velocidad. Con la evolución del sector tecnológico se han introducido dos nuevas características al modelo inicial de las 3 V’s: veracidad y valor. Para entender mejor el fenómeno del *Big Data*, es necesario conocer las propiedades de cada uno de sus cinco atributos:

- *Volumen*: hace referencia a la enorme cantidad de datos que son generados en cuestión de segundos. Esta inmensidad de datos genera dificultades de almacenamiento mediante técnicas de procesamiento tradicionales, lo que impulsa la necesidad de crear nuevos instrumentos y métodos para su administración y análisis de esos datos. Estos instrumentos consisten en infraestructuras adaptables capaces de manejar volúmenes masivos de información, de ahí la creación de grandes centros de datos, repartidos por todo el mundo con gran cantidad de potentes ordenadores, capaces de almacenar y procesar muchos datos.

- *Variedad*: hace referencia a las distintas categorías y tipos de datos en los que se clasifican los volúmenes de datos generados. Predominan los datos estructurados, datos no estructurados y los datos semiestructurados. Esto genera un posible obstáculo debido a que cada tipo de dato cuenta con una estructura que puede dificultar el procesamiento, almacenamiento y análisis de estos. La variedad de datos genera oportunidades para las empresas ya que los datos no estructurados y semiestructurados pueden aportar información importante que no podría obtenerse con tan solo el uso de datos estructurados.
- *Velocidad*: hace alusión a la velocidad en la que los datos son generados, captados, procesados y analizados. El aumento del volumen de los datos al igual que los distintos tipos de datos que existen, hace que la velocidad a la que se realice el tratamiento de los datos deba ser mayor e incluso en tiempo real. Es decir, tiene que existir un ritmo paralelo para poder extraer la información deseada. Nuevas infraestructuras, que hemos mencionado en la primera dimensión, son necesarias para el procesamiento de estos datos que no podrían ser tratados con métodos tradicionales.
- *Veracidad*: la veracidad de los datos consiste en la calidad de los datos obtenidos, que genera fiabilidad en la fuente de la que se han obtenido esos datos. Esta veracidad es fundamental para que el análisis que sea realice, produzca información verídica. Para ello los datos han de ser auténticos, lo que facilitará la correcta interpretación y análisis de la información obtenida a través de esos datos. Tras el aumento en el volumen, variedad y velocidad de datos, es más probable la transmisión de datos erróneos, maliciosos e inconsistentes que pueden dar lugar a una mala interpretación de la información. Por ello, los sistemas de análisis han de dotarse de medios de comprobación de la veracidad de los datos obtenidos (Grover et al., 2018).
- *Valor*: el valor del dato va a ser el componente que engloba las cuatro características explicadas y va a cumplir el objetivo principal del *Big Data*. Al trabajar con grandes volúmenes de datos, se busca encontrar información completa y precisa obtenida de diferentes fuentes de datos. Esta información tiene como propósito crear valor mediante la



contribución de una toma de decisiones más informada y en último término la monetización del dato, aportando dinero a aquellos que sean capaces de obtener y analizar los datos.

Si profundizamos en estos cinco componentes con detalle podemos observar que existe interconexión entre cada una de estas particularidades, haciendo que unas puedan influir en las otras. El volumen de los datos es cada vez mayor, lo que introduce tipos emergentes de datos, que a su vez dificulta el procesamiento de datos a velocidades óptimas. A su vez, comprobar la veracidad de los datos puede afectar a la velocidad a la que se procesan estos datos. Pero el valor va a ser el objetivo final del *Big Data*, aprovechando el volumen, la variedad, la velocidad y la veracidad para obtener información relevante y llevar a cabo decisiones informadas.

Además, es posible que a medida que el *Big Data* y las tecnologías se desarrollen en los próximos años, sea necesario añadir nuevas dimensiones a las 5V existentes, que permitan caracterizar el término *Big Data* de forma más precisa, y así contribuir a la creación de valor en las organizaciones.

### 3.2. Tipos de datos

Para entender mejor los datos que apoyarán la toma de decisiones en las empresas, hay que tener en cuenta la variedad de datos que existen. Cada tipo de dato aporta un valor diferente y único, y por ello las empresas deben desarrollar mecanismos de procesamiento adecuados para explotar el dato de la manera más eficaz, y así poder extraer el conocimiento deseado.

Con el desarrollo de las nuevas plataformas digitales, se está generalizando la disponibilidad de datos que no se pueden estructurar de una manera convencional, y no pueden ser procesados de una manera estructurada. Los datos, actualmente, se encuentran en forma de ficheros de texto, ficheros XML, correos electrónicos, imágenes, *web blogs*, y videos.

Los datos se pueden categorizar en tres grupos (Elia et al., 2020):

- *Datos estructurados*: los datos estructurados son aquellos que siguen una estructura definida. Esto permite un fácil manejo de estos datos utilizando herramientas tradicionales.

La gestión de los datos estructurados se lleva a cabo mediante la utilización de sistemas de bases de datos relacionales, los cuales almacenan números, fechas, cadenas de texto, etc. Estos sistemas incluyen un lenguaje específico de programación denominado SQL, o lenguaje de consulta estructurada (Chen et al., 2009). La analítica de negocios tradicional se ha basado en estos mecanismos de análisis de datos estructurados, relegando el resto de las tipologías de datos.

- *Datos semi-estructurados*: los datos semiestructurados son una mezcla de los datos estructurados y datos no estructurados. Es decir, son datos que están organizados, pero carecen de una estructura. Es más, se podría decir que son datos con una estructura impredecible o cambiante. Los ejemplos más típicos de datos semi-estructurados son (Chen et al., 2009):
  - Correos electrónicos.
  - Ficheros XML (*Extensible Markup Language*). Estos se utilizan para el intercambio de información entre ordenadores, y sus especificaciones se definen dentro del mismo fichero.
  - Ficheros de documentos (Microsoft Word, Google Docs).
  
- *Datos no estructurados*: los datos no estructurados suponen el principal reto para la analítica moderna del dato. Son aquellos que carecen de una estructura, pero que contienen información esencial para su análisis y explotación. Los datos no estructurados predominan en las plataformas digitales modernas, y requieren habilidades específicas para su gestión. Podemos categorizar los datos no estructurados de la manera siguiente (Eberendu, 2016):
  - *Documentos de comunicación*: contenidos de redes sociales (Instagram, Facebook), contenidos web, contenidos de plataformas de mensajería instantánea (Whatsapp).
  - *Documentos dinámicos*: Documentos que pueden ser creados, editados, revisados, aprobados, etc. Ejemplos pueden ser: *White-papers*, artículos, documentos corporativos, presentaciones.
  - *Documentos multimedia*: Ficheros de audio, ficheros de video, imágenes.
  - *Documentos estáticos*: Faxes, documentos en formato pdf.
  - *Contenidos radar*: datos sísmicos, meteorológicos, oceanográficos.

Según Eberendu (2016), los datos no estructurados suponen un 61,8%, mientras que los datos estructurados suponen un 23,7 %. Esto representa un desafío para las empresas ya que las técnicas de análisis de datos tradicionales no se pueden aplicar a los datos no estructurados, requiriendo nuevas herramientas. Pero a su vez, supone una oportunidad para las empresas para descubrir información que aporta perspectivas más amplias.

En el 2020, Zhang et al. desarrollaron un modelo para la predicción en el sector hospitalario, en el que se incluyeron los siguientes elementos a analizar: las estimaciones de mortalidad en el hospital, probabilidad de larga permanencia en el hospital y estimación de readmisión hospitalaria. Para desarrollar estos modelos predictivos, se han utilizado varias categorías de datos. Esta estructura de datos (**Tabla 1**) fue aplicada para el entrenamiento y prueba de un modelo predictivo basado en redes neuronales. Los resultados mostraron que los modelos predictivos que utilizan agregaciones de datos estructurados y no estructurados consiguen mejores resultados.

**Tabla 1: Categorías de datos en modelo predictivo en el sector hospitalario**

<b>Datos estructurados</b>	<b>Datos no estructurados</b> (en textos de notas médicas)
<b>Información del paciente:</b> Información demográfica, edad, sexo, situación marital, etnia, historial clínico.	Notas de visitas de enfermería Notas de admisión hospitalaria
<b>Señales temporales:</b> ritmo cardíaco, presión arterial sistólica y diastólica, ritmo respiratorio, temperatura, así como más de 19 parámetros de análisis clínicos.	Notas de visitas de médicos Notas de radiología y demás pruebas médicas.

### 3.3. El dato y su ciclo

Los datos no son elementos suficientes para poder llevar a cabo un proceso de toma de decisiones. El correcto procesamiento y posterior análisis de los datos va a ser clave para la obtención de información que sustenta la toma de decisiones.

Las organizaciones han de definir claramente un plan para la explotación de los datos, con el objetivo de maximizar su valor para la empresa. Además, las empresas han de distinguir cada paso por el que los datos han de pasar para comprender la extracción del valor, que es el objetivo principal en el ciclo de vida de los datos.

Se ha creado un marco de trabajo, reflejando los pasos que sigue el dato a lo largo de su vida, basado en los pasos propuestos por Stobierski (2021).

#### **Captura**

La captura de datos se divide en tres fases:

- *Generación de datos*: Normalmente la generación de datos se produce gracias a las actividades desarrolladas en internet, tanto en el ámbito de la empresa como en el de las personas. Aunque también puede haber datos que se generen dentro de la propia empresa, de clientes o de proveedores. En internet, cualquier actividad “genera” datos. La clave radica en saber identificar esos datos que pueden aportar valor. Dentro de este paso es donde se debe discernir que tipo de dato es: estructurados, no estructurados o semiestructurados. A su vez, la disponibilidad de los datos es cada vez más abundante, dando lugar a nuevas oportunidades de negocio.

Típicamente, los datos generados en el ámbito de la empresa son gestionados por diferentes aplicaciones corporativas, tales como sistemas ERP (*Enterprise Resource Planners*), CRM (*Customer Relations Management*), etc. Estos datos son fácilmente estructurables (Liu, 2015).

Existen, por ejemplo, relevantes aplicaciones de IoT aplicadas a la agricultura inteligente, abordando un problema tan acuciante como es el crecimiento global de la población, comparado con los recursos naturales. Se presenta un marco de trabajo para el desarrollo de la agricultura inteligente, apoyado en la explotación de los datos generados por dispositivos IoT (Quy et al., 2022).

Otros medios de capturar datos en algunas actividades empresariales es través del IoT (Internet de las Cosas). En el mundo del IoT, cada evento genera un dato. Esto requiere que estos datos generados se envíen a través de protocolos como *MQTT*, *WAMP*, *HTTP* o *Sigfox*. Estos protocolos envían la información desde los dispositivos IoT hasta una base centralizada. En los sistemas IoT cobra especial importancia el “Volumen” y la “Velocidad” del dato (Wytrębowicz, Cabaj y Krawiec, 2021).

Es factible igualmente obtener datos de fuentes externas a la empresa. Puede realizarse mediante interacciones o transacciones como es el caso de las interacciones entre usuarios en las Redes Sociales (Eberendu, 2016).

- *Recolección de datos*: La recolección de datos consiste en la utilización de herramientas para obtener datos específicos, relevantes para los procesos de negocio de la empresa. Se trata de un paso más activo que la generación de datos en sí, ya que se busca obtener información más específica y seleccionar los datos generados que son de interés.

Podemos enumerar diferentes mecanismos de recolección de datos (Hill et al., 2020):

- a. *Forms*: *Google Forms*
  - b. *Surveys*: *Survey Monkey*
  - c. *Medidas de la experiencia del usuario*: Identificando los parámetros fundamentales en la interacción del usuario con el producto o servicio digital
- *Almacenamiento de datos (corto plazo)*: una vez recopilados los datos, se va a proceder al almacenamiento de estos para poder tener fácil acceso a ellos durante su análisis y posterior toma de decisiones.

El almacenamiento de datos estructurados se realiza principalmente en bases de datos relacionales (*DBMS*), de ámbito muy extendido.

Por el contrario, para el almacenamiento de datos no estructurados no es tan trivial se utilizan dos conceptos: *data lakes* y *data warehouses*. Básicamente, estas herramientas permiten el almacenamiento de un volumen masivo de datos, parte de los cuales serán procesados posteriormente. Algunos ejemplos son: *Apache Hadoop*, *NoSQL*, *Apache Hbase*, *Databriks* (Tsai et al., 2022).

### **Preparación**

La preparación de datos consiste en transformar, seleccionar, homogeneizar los datos para que sean utilizables en los pasos siguientes. Este paso va a garantizar la validez en términos de calidad de datos y la fiabilidad a la hora de realizar el análisis de datos. Se utilizan herramientas para hacer la preparación y procesado de los datos que por ejemplo eliminan datos duplicados, normalizan la información, quitan datos erróneos o valores atípicos, etc. Las herramientas comerciales que se utilizan para esta tarea son Python y R, entre otras.

### **Análisis y Visualización**

- *Análisis de datos*: se lleva a cabo una vez que los datos están preparados para su utilización. En este paso se llevará a cabo el modelaje de datos utilizando algoritmos que permiten hacerlo de forma automática. Es una tarea crítica donde se aplican los conocimientos del tema concreto, y que va a permitir obtener información valiosa para la toma de decisiones.
- *Visualización de datos*: en esta fase se define el formato visual que se va a utilizar para presentar la información que mejora la comprensión del análisis que se ha realizado. Mediante la visualización de los datos se facilita la interpretación, comprensión, tendencias y relaciones en los datos.
- *Toma de decisiones*: una vez visualizados y comprendidos los datos que se han analizado, se va a llevar a cabo la toma de decisiones. Se trata del paso clave para la generación de

valor, ya que se va a utilizar la información obtenida a través del análisis de los datos, para elegir la mejor opción entre varias alternativas. A continuación, se implementa la decisión tomada, y posteriormente se ha de realizar un seguimiento de esta para ver si se ha cumplido el objetivo que había.

Las herramientas que se utilizan en esta fase de análisis y visualización son las siguientes (Diamond y Mattia, 2017):

- *PowerBI/Excel*
- *Tableau*
- *Jupyter Notebook*
- *Apache Spark*
- *SAS*
- *Knime*

Cabe resaltar la potencia de Python, como entorno de desarrollo de código abierto, donde existen paquetes de adquisición de datos, procesamiento, e incluso visualización, que permiten desarrollar sistemas comerciales de explotación de datos (Yu y Yuan, 2022).

### **Gobierno**

El gobierno del dato, en general, es un tema complejo que requiere especial atención en el mundo empresarial. Se trata del conjunto de procesos encaminados a garantizar la integridad, la seguridad, la veracidad y la administración en general de los datos de una empresa (Petzold et al., 2020).

Si el gobierno del dato no está garantizado, entonces las decisiones basadas en los datos serán erróneas. Por tanto, este elemento es un pilar fundamental en la estrategia de *Big Data* de una compañía. Todos estos procesos deben estar garantizados por un equipo de gestión del dato, encabezado por un director de datos.

### 3.4. Marco de clasificación del Big Data

Previamente se ha descrito el ciclo de vida del dato, mediante el cual se identificó el tratamiento de los datos desde su generación hasta el almacenamiento de estos. Para entender mejor el ciclo de vida del dato, hay que seguir un proceso que guiará la creación del valor. El punto de partida para crear valor con *Big Data* es contar con infraestructura donde se realicen todos los pasos del ciclo que siguen los datos. Se debe contar además con unos mecanismos y capacidades, que establezcan unos objetivos a alcanzar y que una vez implementado, pueda medirse el impacto. Cada uno de los siguientes elementos se desarrollan y establecen acorde con el camino y sentido que quiere tomar la empresa (Grover et al., 2018).

#### 3.4.1. Elementos esenciales en Big Data para la creación de valor

**Infraestructura BDA:** la inversión en la infraestructura del *Big Data* va a ser determinante para los siguientes pasos en el proceso de tratamiento de datos. La empresa debe enfocar la inversión de sistemas de tal manera que las herramientas estén alineadas con los futuros procedimientos y decisiones que se lleve a cabo. Esta inversión se considera parte de los activos de la empresa y del *portfolio* de las herramientas de análisis, mediante los cuales se extraerá valor e información a partir de los datos. También cabe destacar el papel fundamental que toma el talento humano, que es necesario para llevar a cabo el tratamiento de los datos y su extracción posterior de valor. Sin profesionales que sepan manejar estas infraestructuras no sería posible el tratamiento del *Big Data* y su posterior interpretación al igual que la extracción de información.

**Capacidades BDA:** Para aprovechar al máximo el BDA, las empresas deben tener capacidades en áreas como análisis de texto, análisis predictivo, análisis de audio y video, análisis de redes sociales, análisis geográfico, análisis en tiempo real y análisis de grafos. Estas capacidades les permiten obtener información valiosa de los datos y tomar decisiones basadas en *insights* en tiempo real.

**Mecanismos de creación de valor:** El análisis de datos empresariales tiene diferentes formas de crear valor. Algunos de los mecanismos clave incluyen la transparencia y el acceso a la



información, el descubrimiento y la experimentación, la predicción y optimización, la personalización de productos y servicios, el aprendizaje y el crowdsourcing, y la capacidad de monitoreo y adaptación. Estos mecanismos permiten tomar decisiones más informadas, segmentar clientes de manera más precisa, mejorar el rendimiento y la experiencia del cliente, desarrollar productos y servicios inteligentes, y prevenir problemas futuros.

**Objetivos de valor:** las empresas que adaptan mecanismos tecnológicos en la actividad empresarial y operativa tienen como finalidad generar valor en algún punto de la actividad empresarial. Las empresas son cada vez más ambiciosas y buscan adoptar la mayor cantidad de medidas posibles para poder optimizar los procesos operativos. Pero para poder llevar a cabo una estrategia empresarial basada en el *Big Data*, es necesario tener un objetivo claro y evitar un enfoque no dirigido. Los objetivos de la organización que implementan *Big Data* suelen estar relacionados con el rendimiento de los empleados y la organización global, la reducción en los costes operativos, progreso en los procesos de producción y la mejora de productos y servicios para perfeccionar la experiencia de los clientes (Raguseo, 2018). Estos objetivos están alineados con el marco de clasificación desarrollado del *valor según su dirección*, que se encuentra explicado más adelante.

**Impacto:** es el momento más importante en la creación de valor de los datos, ya que va a determinar la efectividad del resto de los elementos. Es decir, el impacto mide si las infraestructuras establecidas han sido las adecuadas en base al sentido que quiere tomar la empresa, y si las capacidades han extraído la información que la empresa busca con respecto a los objetivos de creación de valor.

## 4. El valor estratégico del dato

### 4.1. El concepto del valor

El valor podría definirse como una aptitud que aporta significado, ya sea a una persona o a una cuestión. Cuando añadimos valor a algo, lo hacemos máspreciado. Pero el valor no se genera de forma fácil, sino que lleva consigo un proceso lógico.

Pero nos interesa analizar el valor en un ámbito empresarial. En este entorno, el valor consiste en aportar una ventaja competitiva a un producto, un bien o servicio, o incluso a una empresa si se lleva a cabo una estrategia empresarial adecuada. Pero para llevar a cabo estas transformaciones en el modelo de negocio, hay que tener claro los objetivos que la empresa trata de alcanzar. El valor es un componente que toda empresa ha de buscar, ya que, si se trabaja en una estrategia empresarial para conseguir ese valor buscado, se generará crecimiento empresarial. Más concretamente, en la era digital en la que nos encontramos, el valor empresarial es generado a través de los avances tecnológicos que se implementan en las organizaciones. Estas nuevas tecnologías, como el *Big Data* permiten sacar el máximo partido a los datos, y por consiguiente generar valor (Elia et al., 2020). El valor es el objetivo principal a la hora de aplicar tecnología a los datos para descubrir información oculta que puede ser útil para la elaboración de estrategias empresariales.

Dado que la importancia del *Big Data* y las nuevas tecnologías es cada vez mayor, adoptar marcos estratégicos digitales que puedan localizar fuentes de creación de valor es crucial para las empresas. Según Yoo, Henfridsson, & Lyytinen (2010), la capacidad de crear nuevos productos, la heterogeneidad, las plataformas de productos digitales y la capacidad de creación de significado son fuentes de creación de valor a través del *Big Data*.

En primer lugar, se encuentra la capacidad de crear nuevos productos (*generativity*). Por ejemplo, las diversas oportunidades que tienen los usuarios hoy en día de acceso a herramientas y recursos necesarios para crear nuevo contenido.

En segundo lugar, la heterogeneidad tecnológica, que permite el desarrollo de bienes y servicios personalizados dirigidos a diversos segmentos de clientes. Esto hace posible el desarrollo de tecnologías adaptadas a las distintas demandas y preferencias de los consumidores, lo que se traducirá en una ventaja competitiva y en mayores niveles de satisfacción general de los clientes.

Por otro lado, las infraestructuras tecnológicas, conocidas como plataformas de productos digitales, que permiten el desarrollo y la distribución de bienes y servicios digitales.

Por último, la capacidad de creación de significado, que permite aprovechar las tecnologías para establecer conexiones con los clientes y proporcionar un valor que vaya más allá de la funcionalidad básica del bien o servicio. Dicho de otro modo, tener la capacidad de crear significado implica proyectar y ofrecer experiencias que vayan más allá de satisfacer las demandas fundamentales e iniciales de los clientes, para así aumentar su fidelidad y mejorar la percepción de la marca.

#### 4.2. El beneficio económico del valor

En la mayoría de los casos, cuando se habla de valor se relaciona con los beneficios económicos. Por tanto, cuando se hace una transformación en el modelo de negocio de las empresas, se puede crear valor empresarial, que posteriormente se traducen en beneficios económicos (Elia et al., 2020).

El análisis de datos permite llevar a cabo decisiones más informadas con respecto a los procesos empresariales. Esto se traduce en extraer información que va a ayudar a la empresa a conseguir una reducción de costes y a una mayor eficiencia operativa. También se pueden identificar nuevas oportunidades enfocadas al crecimiento del negocio.

Es más, según Chen, Preston y Swink (2015), estos beneficios esperados obtenidos gracias al BDA se clasifican en dos categorías:

- **Beneficios directos:** estos beneficios van a estar relacionados con la reducción de costes y mejora en eficiencia.
- **Beneficios indirectos:** estos beneficios están relacionados con las emergentes oportunidades que ofrece el *Big Data*. Es el caso de la mejora del servicio al cliente.

Pero, en definitiva, sin importar que sean beneficios directos o indirectos, se traducen en beneficios económicos. Es decir, la mejora del servicio al cliente va a generar un entorno de confianza por parte del consumidor que hará que en un futuro vuelva a realizar una compra en la misma empresa.

### 4.3. Captura de valor y creación de valor

En el contexto del valor, es muy importante tener en cuenta la diferencia entre la creación de valor y la captura de valor, ya que no significa lo mismo.

En el contexto de *Big Data*, la creación de valor implica aprovechar los datos para generar ventajas competitivas, mientras que la captura de valor se refiere a la capacidad de traducir ese valor en beneficios económicos tangibles. De nada sirve crear valor si no se lleva a cabo una estrategia empresarial en la que no se implemente el valor extraído. Ambos aspectos son fundamentales para maximizar el potencial de *Big Data* y obtener un retorno significativo de las inversiones realizadas (Urbinati et al., 2019).

El uso de la tecnología tiene como objetivo maximizar los recursos y la productividad. Por otra parte, el valor se crea cuando los beneficios superan los costes para obtener ese valor. Es por ello por lo que con la utilización de *Big Data* se desarrollan nuevas herramientas y métodos para contribuir a esa creación de valor.

Según Elia et al. (2020), muchas empresas no llevan a cabo la implementación de *Big Data* en su actividad empresarial, impidiendo la creación y captura de valor que se puede generar con estas herramientas tecnológicas, porque no perciben el valor que las aplicaciones tecnológicas pueden ofrecer o porque no están preparadas para dar ese paso ya sea por falta de recursos o conocimientos.

Existen empresas que proporcionan soluciones basadas en *Big Data*, que se dedican a exponer el valor que puede ofrecer la implementación tecnológica en las estrategias empresariales. Estas empresas ayudan a sacar a la luz este conocimiento que el *Big Data* aporta, pero es traducido en valor para la empresa. También ayudan a no dejar escapar ese valor, es decir, a capturar ese valor, mediante soluciones que proponen a la empresa (Elia et al., 2020).

Wheeler (2016) argumenta la importancia de comprender el valor como primer paso para la creación y captura del mismo. Sin este conocimiento del valor, no se va a poder hacer uso de estas herramientas para conseguir los objetivos de la empresa.

#### 4.4. Marco de clasificación del valor

Según Rajpurohit (2013), el establecimiento de un marco de referencia para la creación de valor es imprescindible para poder ejecutar proyectos con *Big Data*. Existen numerosas oportunidades y aportaciones del *Big Data* en las empresas, pero sin ese marco de referencia podría surgir una brecha entre el valor real y el valor potencial. Elia et al (2020) propone un marco de clasificación del valor según su dirección, que contribuye a la comprensión de la dirección que quiere tomar una empresa con respecto de sus objetivos y misión del futuro. Este marco de clasificación presenta nueve direcciones diferentes que van a apoyar la toma de decisiones empresarial a partir de información extraída con técnicas de análisis de datos.

##### 4.4.1. El valor según su dirección

**Posicionamiento en el mercado:** el posicionamiento de mercado se trata de una dirección de valor muy significativa que puede aportar el *Big Data*, de forma que las empresas pueden extraer información acerca de cómo posicionar los productos y servicios que ofrecen de tal manera que estén dirigidos al mercado deseado. Un bien o servicio no es significativo si no está dirigido al público correcto, ya que la estrategia empresarial no resultará efectiva ni producirá los resultados esperados.

El posicionamiento en el mercado puede ser ajustado y conseguido alrededor de las 4 P's: producto, precio, promoción y punto de venta (Erevelles, Fukawa, y Swayne, 2016).

A través del *Big Data* se puede realizar un seguimiento de tendencias del público objetivo y determinar el precio que estos clientes estarían dispuestos a pagar y así poder crear oportunidades de diferenciación de precios.

El *Big Data* también contribuye al desarrollo innovativo de nuevos productos y servicios o incluso mejoras en base a la demanda que observan en los consumidores, sin tener que depender de técnicas tradicionales de investigación de mercado.

Por otra parte, el *Big Data* también proporciona información y perspectivas en torno a las preferencias de los clientes y sus comportamientos y así poder establecer campañas de marketing adecuadas para poder lanzar el mensaje correcto en el momento adecuado y llegando al público indicado.

Por último, cabe destacar el rol del *Big Data* en la distribución de productos y servicios para el posicionamiento adecuado en el mercado. Los datos ayudan a las empresas a determinar los puntos de distribución más adecuados para garantizar el acceso de los bienes o servicios en el mercado objetivo.

**Capacidad de respuesta del mercado:** una vez la empresa tenga claro el mercado al que se quieren dirigir, es necesario mediar la respuesta que está teniendo para analizar si las expectativas se alinean a la realidad. Por ello, el *Big Data* servirá como herramienta para facilitar soluciones a los cambios inesperados en el mercado.

En este ámbito, el *Big Data* juega un papel fundamental al poder realizar un análisis del mercado con información a tiempo real. Estos datos extraídos contribuyen al estudio de las tendencias y patrones no solo de los competidores en el mercado, sino que también de los consumidores y la demanda.

**Fidelización de la clientela:** los datos son una fuente de información muy valiosa en cuanto a perspectivas acerca de los consumidores. El análisis de datos puede ayudar a crear una base de clientela leal proporcionando información acerca de los comportamientos y preferencias de los clientes para así crear la mejor experiencia del cliente posible que pueda resonar en el futuro.

Al obtener información acerca de los clientes, las empresas son capaces de generar valor creando experiencias personalizadas y también crear campañas de marketing más dirigidas. La información generada también puede tener forma de comentarios y *feedback* para identificar las áreas de mejora, creando cercanía con los clientes y por consiguiente un aumento de fidelidad hacia la organización.

**Desarrollo de competencias:** con el surgimiento del *Big Data*, nuevos puestos de trabajo se han establecido y a su vez, nuevas competencias son requeridas. Para que las empresas sean capaces de implementar estrategias empresariales, es necesario emplear a personal que pueda usar las herramientas para manipular los datos. Pero el *Big Data* también proporciona conocimientos que contribuyen con el desarrollo de estas competencias.

Por ejemplo, cada vez es más común el aprendizaje de lenguajes de programación y análisis predictivo para muchos trabajos que incluso no tienen relación directa con los puestos de científicos de datos. Estos conocimientos y competencias pueden ser adquiridas por lo tanto a través de las emergentes herramientas de *Big Data* (Lamba y Dubey, 2015).

**Aumento de los ingresos:** el *Big Data* juega un papel muy importante en los beneficios e ingresos de las empresas. Cabe destacar la relación que este componente muestra con los clientes, ya que el análisis de datos permite incluso prever las posibles pérdidas de clientes que a su vez va de la mano de la pérdida de ingresos (Arora y Malik, 2015).

Mediante un análisis de datos las empresas pueden incluso buscar maneras de mejorar los bienes o servicios para explotar información que originalmente se encuentra oculta a través de técnicas de análisis de datos con el objetivo de mejorar los ingresos.

**Aumento de la productividad:** la mejora en la productividad empresarial es una de las principales direcciones que ofrece el *Big Data* en las empresas. Esto es alcanzado a través de la optimización de procesos de producción y distribución.

Por otra parte, las nuevas tecnologías permiten realizar seguimientos del desempeño de los empleados de una empresa, lo que permite obtener información sobre sus fortalezas y debilidades. Esta información puede ir dirigida a un plan de desarrollo de competencias mediante el cual se ofrece a los empleados con el entrenamiento necesario para mejorar su rendimiento en la empresa.

**Ahorro de costes:** una de las ventajas que muchas empresas encuentran en las nuevas tecnologías es la capacidad de reducir costes tras la implementación de avances en sus procesos productivos. El *Big Data* permite identificar ineficiencias a tiempo en los procesos de la empresa a través del análisis de los datos que suele tener un gran volumen. Este análisis aporta perspectivas útiles relacionadas con la asignación de recursos, permitiendo así la reubicación de estos para optimizar los medios que posee la empresa.

Principalmente, la reducción de costes en las empresas puede ser alcanzada a través de la optimización conseguida con el *Big Data* en la cadena de suministro. El análisis de datos en este ámbito permite realizar modificaciones en los niveles de inventarios, proveedores e incluso transporte.

Por último, el Big Data ayuda a las empresas a identificar y mitigar riesgos antes de producirse en las actividades empresariales, evitando pérdidas monetarias.

**Descubrimiento de conocimientos:** el descubrimiento de información y conocimientos está estrechamente ligado a la utilización de los datos como apoyo en la toma de decisiones. Estos conocimientos e información suelen estar ocultos a simple vista y por lo tanto se requieren de técnicas análisis de datos para poder sacar a la luz esta información. Las técnicas de análisis de Big Data manejan grandes volúmenes de datos de diversas fuentes, lo que permite realizar un análisis más completo al incorporar datos de distintos orígenes.

Estas herramientas también permiten la identificación de tendencias o posibles evoluciones que no pueden ser posibles mediante el uso de técnicas de análisis tradicionales. Estas tendencias son creadas a partir de la información extraída de los datos que han sido analizados y que estaban inicialmente escondidos o incluso de datos a tiempo real. Los conocimientos extraídos a tiempo real van a ayudar en la toma de decisiones en ambientes dinámicos.

**Mejora de la infraestructura informática:** las tecnologías del *Big Data* permiten procesar datos a mayor escala, proporcionando una superior maximización de la infraestructura informática de la



empresa. Esta extensa infraestructura informática permite realizar un análisis más complejo mejorando el proceso de toma de decisiones.

A su vez, el *Big Data* también permite realizar un seguimiento más adecuado de sus equipos de *IT* para asegurar el correcto funcionamiento de los sistemas informáticos mejorando su fiabilidad.

#### 4.5. Limitaciones en la creación de valor

A pesar de que las tecnologías tienden a contribuir a la transformación empresarial que han experimentado muchas industrias y negocios, actualmente la implantación de soluciones tecnológicas y BDA plantean una serie de limitaciones, derivadas de los retos en la ejecución de estos proyectos.

Se ha demostrado que una ejecución exitosa de proyectos que incluyen BDA, aporta valor a la empresa. Ahora bien, los retos y fracasos de las soluciones BDA restan valor a la empresa. Un proyecto de BDA no es trivial, sino que es complejo y caro. En muchos casos, los proyectos de BDA terminan en fracaso, suponiendo un gran esfuerzo económico. De hecho, según Ermakova et al. (2021), el 87% de los proyectos de datos no llegan a fase de producción.

A continuación, se describen una serie de limitaciones en la creación de valor a través del Big Data y las nuevas tecnologías.

##### 4.5.1. El coste de la infraestructura

Es indudable que el almacenamiento y procesamiento de datos tiene un coste alto. A la hora de desarrollar un proyecto de BDA, deben tenerse en cuenta los costes asociados a la infraestructura requerida, y al procesamiento de grandes cantidades de datos.

Es el caso concreto de almacenamiento y computación en biología la secuenciación de genes es un proceso que maneja gran cantidad de datos. Requiere además mecanismos de interconexión entre centros de datos de alta velocidad y baja latencia (Marx, 2013).

Por lo tanto, estos elevados costes dirigidos a las infraestructuras generan dudas e incertidumbre en la rentabilidad de la inversión. Por lo tanto, el coste de la infraestructura debe considerarse como un elemento importante a la hora de determinar la distribución de recursos para alcanzar los propósitos empresariales.

#### 4.5.2. Los datos no estructurados

Este trabajo ha remarcado la importancia de los datos no estructurados, como un componente primordial para la obtención de información valiosa. El problema no es trivial, puesto que la extracción de información sobre datos no estructurados requiere un tratamiento más complejo y caro que en el caso de los datos estructurados.

A continuación, se enumeran una serie de ejemplos del complejo tratamiento de los datos no estructurados (Tanwar, Duggal y Khatri, 2015):

- *Analítica de texto*: Se trata de extraer información de piezas de texto como correos electrónicos, blogs, anuncios, redes sociales, transcripciones de centros de atención de llamadas, etc. Para ello, se requiere la aplicación de técnicas de *Natural Language Processing*, entre otras.
- *Analítica de video*: Esta disciplina requiere el manejo de grandes cantidades de datos de video, considerando un video como una secuencia de campos, típicamente 50 campos por segundo. En esta área, tendremos que aplicar técnicas de *Computer Vision*, que caen bajo la disciplina de la inteligencia artificial.
- *Analítica de las redes sociales*: Esta analítica requiere competencias de psicología, sociología, antropología, patrones de comportamiento, etc. Se trata de identificar tendencias, opiniones, sentimientos, y conclusiones muy valiosas, que no son triviales en su extracción.

#### 4.5.3. Protección y privacidad de los datos

Sin duda, uno de los principales retos a los que las nuevas tecnologías se enfrentan hoy en día es la privacidad de los datos personales y su seguridad. Existen diversas causas por las que la privacidad de los datos supone una amenaza.

En primer lugar, cabe destacar que a medida que se crean nuevas técnicas de recopilación de datos, se mejora la capacidad de obtención de información. Los avances tecnológicos han permitido recopilar datos en tiempo real, de las personas que utilizan dispositivos personales (smartphones, relojes, etc).

Según Merchant et al. (2021), se ha demostrado que mediante la actividad de usuarios en las redes sociales es posible predecir enfermedades de estos usuarios mediante datos recogidos en estas plataformas. Es más, se construye un modelo predictivo dando lugar a información acerca de los usuarios usando información confidencial, que ha sido revelada por medios ajenos a los propios usuarios. Esto supone un dilema, que pone en compromiso a las organizaciones como las redes sociales al no tener un control de la privacidad de los datos generados en sus plataformas.

Muchos estarán de acuerdo con que en esta situación está generando valor al proporcionar un diagnóstico sobre una enfermedad con la simple recolección de datos, pero ¿dónde se encuentra en el consentimiento por parte de los usuarios y por consiguiente su privacidad? Esta preocupación ha generado la necesidad de una regulación específica sobre la protección de los datos de la ciudadanía.

En 2018 entró en vigor la Ley General de Protección de Datos. Esta ley establece el modo en que las empresas deben recopilar y procesar la información que mantienen de los ciudadanos, ya que tenemos derecho a la protección de nuestros datos personales (Regulación, 2018).

Muchas organizaciones también llevan a cabo labores de marketing mediante las cuales los consumidores reciben recomendaciones personalizadas acerca de productos y servicios. Muchos consumidores pueden valorar positivamente estas recomendaciones personalizadas, pero otros

muchos lo consideran una intromisión a su privacidad, y sienten preocupación ante la exposición de información personal y la explotación de los datos por parte de las organizaciones. Esto a su vez puede fomentar el rechazo a la publicidad y estrategias de marketing que llevan a cabo las empresas. ¿Se trata de una apropiada estrategia empresarial obtener información personal de los usuarios? Esta es una pregunta que las organizaciones han de preguntarse ante este emergente reto (Shamsuzzoha y Raappana, 2021).

A pesar de que muchos usuarios tratan de cuidar su privacidad, existe una creciente tendencia a vincular muchos aspectos de sus vidas con las tecnologías. En el caso de las redes sociales, los usuarios pueden aportar datos de gran valor a las empresas como, por ejemplo, ubicación, opiniones políticas, gustos en cuanto a ropa, comida, etc.

Por otro lado, nos encontramos con el caso del consentimiento de las cookies o términos y condiciones cuando los usuarios acceden a una página web. En un estudio de evaluación de usabilidad de las interfaces de consentimiento de cookies realizado, se descubrió que el 50,2% de los participantes aceptan todas las cookies con el objetivo de deshacerse cuanto antes de la interfaz de consentimiento. Definitivamente, aquellos que aceptan el consentimiento para poder empezar su navegación en la página web no leen detalladamente el contenido de este consentimiento. Es por ello, que se puede pensar que la mayoría de la gente pasa por alto su seguridad, ya que lo perciben como una molestia cuando ven aparecer esa interfaz (Habib et al., 2022).

De las personas las cookies, el 57,9% revelaron que solo aceptan las cookies que son estrictamente necesarias, lo hacen con el objetivo de mantener su privacidad y así limitar la exposición de información personal que puede ser utilizada para diferentes fines por parte de las empresas. Es decir, muchos usuarios son conscientes de que las empresas pueden obtener valor a partir de los datos que generan los usuarios (Habib et al., 2022).

Además, estar envuelto en un escándalo de privacidad de datos puede suponer un inconveniente para la empresa, debido al daño que puede causar en la reputación de la organización y por consiguiente evitar la llegada de nuevos clientes. Por tanto, las empresas tienen que poner sobre la

balanza si deben priorizar la seguridad de sus usuarios o si prefieren obtener información que consideran de valor para la toma de decisiones que han de llevar a cabo con respecto a sus clientes.

En conclusión, la conectividad se ha convertido en una necesidad básica y la globalización ha fortalecido las conexiones interpersonales. Sin embargo, esta dependencia puede comprometer la privacidad y seguridad de los usuarios. Aunque la protección de datos y la preocupación por la privacidad están aumentando, muchos usuarios aún encuentran difícil abandonar ciertos servicios debido a su arraigo en la vida cotidiana. Las empresas que dependen de la confianza de los clientes se enfrentan a amenazas serias si no abordan adecuadamente estas preocupaciones, ya que pueden perder clientela y dañar su reputación. Para mantener la confianza de los usuarios, las empresas deben tomar medidas estrictas en cuanto a la privacidad y seguridad de los datos.

#### 4.5.4. Escasez de competencias y talento

Wu, Liang y Chen (2022) plantean una problemática muy representativa para la extensión de la utilización del Big Data y es la falta de competencias de gestión de datos, y la asignación de recursos en la empresa.

Además, las competencias digitales necesarias en la empresa están distribuidas en diferentes áreas, complicando la obtención de una visión holística del Big Data. Pero para realizar esta gestión integral se requieren perfiles con alta capacidad de gestión y visión estratégica de la empresa, que sean capaces de resolver retos digitales y analíticos, y aplicarlos a la creación de valor empresarial.

## 5. Caso real: Duolingo y reCAPTCHA

### 5.1. Introducción

Duolingo es una plataforma de aprendizaje de idiomas online que fue creada en el año 2011 por Luis von Ahn y por Severin Hacker. Su creador, Luis von Ahn, tenía como objetivo crear una plataforma gratuita para que todas las personas tengan las mismas oportunidades a la hora de tener acceso a recursos de aprendizaje de idiomas, sin importar su poder adquisitivo. Según la entrevista que ofreció Luis un año después del lanzamiento de la plataforma, consideró que la educación genera desigualdad, ya que solo aquellos que tienen la capacidad de financiar los costes académicos pueden tener acceso a programas educativos. Luis es procedente de Guatemala y pudo observar cómo la gente de su país luchaba para salir de la pobreza. En cambio, se dio cuenta de la imposibilidad de salir de la pobreza cuando no se cuenta con los recursos necesarios para invertir en educación. Es por ello por lo que Luis pensó en la financiación como solución a la pobreza y para un aumento de crecimiento social en Latinoamérica, ya que los servicios ya existentes requerían del pago de una suscripción (TED-Ed Español, 2016, 2m19s).

La idea inicial de Luis era crear un servicio gratuito para que los usuarios pudiesen disfrutar de algo tan cotidiano como ir al gimnasio, pero que muchos no pueden realizar al no tener el poder adquisitivo para adquirir una membresía. Pero como toda empresa, hace falta generar beneficios para poder hacer frente a todos los gastos que una organización conlleva. Es por eso por lo que Luis von Ahn tuvo que pensar en una forma de generar beneficios sin cobrar a sus usuarios y se le ocurrió que los mismos usuarios del gimnasio podrían crear electricidad a través de la energía cinética. Es decir, los usuarios crearían valor para la empresa en forma de energía a través del deporte que realizan los usuarios y así vender esa energía posteriormente otras empresas. En cambio, esta idea no pudo llevarse a cabo ya que por lo general los gimnasios crean beneficios cuando sus clientes no acuden al gimnasio, ya que cobran la membresía, pero en esta situación no se genera energía. Esto no era compatible con la idea inicial que Luis quería llevar a cabo, ya que, para poder implementar este plan, era necesario que los clientes acudiesen al gimnasio (TED-Ed Español, 2016, 0m14s).

Luis decidió reciclar su idea inicial de ofrecer un servicio gratuito, pero utilizando un enfoque diferente: la educación. Así nació Duolingo, ofreciendo aprendizaje gratuito a sus usuarios y a su vez aprovechando la motivación de las personas para crear valor a través de la traducción de artículos y libros. Los alumnos no serán cobrados por llevar a cabo estas traducciones ya que practican el idioma que tratan de aprender y, además, tienen acceso a una plataforma de aprendizaje gratuito. A su vez Duolingo usará estas traducciones para generar ingresos, ya que son solicitadas por las empresas de las noticias y libros. Para asegurarse de la generación de valor de calidad, se recopilan las traducciones que han generado los usuarios, escogiendo aquella que mayor valoración reciba. Una vez realizada la traducción completa, el solicitante de dicha traducción pagará a Duolingo por el servicio realizado por los clientes (Gavarri, 2016).

Luis es el cofundador de reCAPTCHA, también conocido como “*Completely Automated Public Turing test to tell Computers and Humans Apart*”. Se trata de un sistema mediante el cual se verifica la utilización de sitios webs por seres humanos, evitando los denominados *bots*. reCAPTCHA se encarga de generar una serie de letras y números creando una combinación única, que mezclado con un fondo distorsionado se genera un código para que el usuario que quiera acceder a la página web lo descifre y verifique que se trata de un ser humano (Ahn et al., 2008). Se trata de una herramienta utilizada mundialmente. De hecho, Luis desveló en su entrevista que se generan alrededor de 200.000.000 códigos al día. Pero la pregunta es la siguiente: ¿con qué fin se creó esta iniciativa? (Benny Lewis, 2012).

Al igual que con Duolingo, reCAPTCHA no solo cumple con la función percibida por el usuario, sino que existe un aprovechamiento secundario. Esta herramienta permite la verificación de los usuarios como seres humanos y a su vez está ayudando a digitalizar escritos que han sido escaneados, pero no reconocidos por el reconocimiento óptico de caracteres (Von et al. 2008).

Ahn (2012) hace referencia al término *eficiencia social*, haciendo alusión a la utilización de recursos para satisfacer necesidades sociales. En este aspecto, Duolingo se encuentra un paso más adelante que reCAPTCHA, ya que los recursos son utilizados para proporcionar conocimientos a los usuarios de manera gratuita. En cambio, reCAPTCHA no proporciona ningún beneficio específico a los usuarios, nada más que el acceso al sitio web y verificar que son humanos. Es

decir, en este caso, el beneficio lo va a tener reCAPTCHA y su cliente, que será el propietario de la página web, ya que obtienen tanto la traducción de caracteres como la verificación de los usuarios.

En definitiva, tanto reCAPTCHA como Duolingo tienen como objetivo sacar valor de los usuarios a través de la asignación de una “tarea”. En el caso de Duolingo, no solo ofrecen material de aprendizaje a estos alumnos como otras empresas competidores en el mercado, sino que además mejoran el proceso de traducción de los textos digitales que se encuentran en internet. Por otro lado, reCAPTCHA ofrece a las empresas la validación de los seres humanos como usuarios de las páginas web con un propósito de seguridad y a su vez se benefician de la interpretación de los códigos que proporcionan para mejorar el algoritmo de reconocimiento de caracteres. Es decir, en ambos casos, las empresas usan los datos que han generado los usuarios como activo, ofreciendo información valiosa para las empresas.

## 5.2. Análisis empresarial

Duolingo ha creado un modelo de negocio en el que todas las partes interesadas e involucradas en el negocio extraen valor. Por una parte, los consumidores se benefician de los conocimientos lingüísticos de la plataforma. Por otra parte, los solicitantes de traducciones se benefician de las traducciones de libros y textos, que a su vez pagan por el servicio. Por otro lado, Duolingo se beneficia al cobrar este servicio a los solicitantes, que a su vez ha sido elaborado por los clientes. Y todo esto es posible gracias a la interconectividad tecnológica y la generación de datos entre las diferentes partes interesadas, extrayendo valor en tres direcciones diferentes. Pero para entender cómo Duolingo y reCAPTCHA generan valor a través del dato, es necesario llevar a cabo un análisis empresarial del valor según su dirección.

En primer lugar, Duolingo tiene un enfoque muy claro con respecto al aprovechamiento del valor a partir de los datos extraídos. Se trata de una plataforma en la que los usuarios interactúan constantemente con el servicio, permitiendo a Duolingo analizar y recopilar gran cantidad de información acerca del progreso del alumno. Duolingo también extrae información acerca de los patrones de aprendizaje de los alumnos. A su vez, Duolingo se encarga de utilizar esta información



que ha conseguido extraer para mejorar la experiencia de aprendizaje del usuario y así poder desarrollar nuevas lecciones o incluso nuevos cursos que garantice la permanencia del alumno en la plataforma. Esto ayuda a Duolingo a entender qué tipo de aprendizaje buscan sus clientes y cómo posicionar la marca en el mercado. En este caso, Duolingo se dio cuenta de que no existía un servicio gratuito de aprendizaje de idiomas y por ello decidió posicionar la plataforma en el mercado como un servicio gratuito y accesible para todo tipo de usuarios.

Por otra parte, Duolingo demuestra su capacidad de respuesta del mercado mediante la mejora en el contenido de sus cursos y la personalización de la experiencia de cada alumno a través de la creación de algoritmos de aprendizaje y de sistemas de recomendación. Todo esto se ha conseguido gracias a la recopilación de datos sobre los usuarios y sobre las tendencias del mercado. Duolingo es una plataforma diseñada para su uso diario, permitiendo así poder analizar información de los usuarios a tiempo real. Es más, en cada lección Duolingo repite aquellas palabras y frases que no han sido respondidas correctamente por los usuarios. Así, Duolingo se asegura de que los usuarios comprenden el material que están aprendiendo, ya que va a ser necesario para realizar más adelante las traducciones de textos y libros que al fin y al cabo aportan beneficios a la empresa.

En cuanto a la fidelización, Duolingo extrae valor de los datos que recopila de los usuarios para crear elementos de fidelización en la plataforma como es el caso de los logros que se pueden obtener tras un uso constante y comprometido de la plataforma. Esto garantiza la fidelización de los usuarios más comprometidos a largo plazo ya que se verán motivados a conseguir los logros y metas personales. Además, los usuarios de Duolingo tienen la ventaja de formar parte en la toma de decisiones de su experiencia de aprendizaje. Por ejemplo, los usuarios son los que van a elegir cuanto tiempo quieren dedicar al estudio diario. Esto hace que los usuarios marquen sus propios objetivos personales y sea más fácil mantener la fidelidad con el servicio. Hay que tener en cuenta que la fidelización también se produce hacia los clientes, que en este caso son las empresas solicitantes de traducciones. Según Hacker (2014), un traductor profesional es capaz de traducir alrededor de 2.500 palabras al día y el coste oscila entre 5 y 35 centavos por palabra dependiendo de diversos factores como la calidad. Esto supone un coste total de entre \$125 y \$875. Si partimos de la base que los usuarios de Duolingo no son pagados por realizar estas traducciones, los clientes

de Duolingo se están ahorrando por su parte mano de obra. Esto genera una fidelización de los clientes al tratarse de una alternativa más económica a los servicios de traducción profesional.

Pero para hacer posible este análisis de algoritmos para mejorar la experiencia de los usuarios, ha sido necesario incorporar nuevas posiciones en la empresa que hagan posible el uso de las herramientas de *Big Data*. Es el caso de la traducción de textos ejercida por los usuarios, mediante la cual el personal informático de Duolingo tiene que asegurarse de que se están llevando a cabo con éxito y que la combinación de las distintas traducciones genera la respuesta óptima. Duolingo cuenta con un sistema de elección de la traducción más próxima al significado real de aquello que se está traduciendo. Es por ello por lo que los empleados de Duolingo han de desarrollar competencias técnicas para poder regular estos procesos de control.

Duolingo ofrece una plataforma gratuita para sus clientes, pero sigue necesitando de una fuente de ingresos para hacer frente a todos los gastos de la empresa. Es por ello por lo que, a través del *Big Data*, Duolingo ha sido capaz de monetizar la plataforma al combinar la necesidad de traducción de textos y libros por parte de las empresas y el deseo de los usuarios de aprender nuevos idiomas. Además de los ingresos recibidos por parte de las empresas solicitantes de traducciones, Duolingo también ha creado una versión premium denominada *Duolingo Plus* mediante la cual los usuarios pueden acceder a la plataforma sin necesidad de conexión a internet. Otra fuente de ingresos de Duolingo son los *English Test*, que permite a sus usuarios realizar un examen de nivel en inglés online, a un precio asequible con respecto a otros exámenes alternativos. Estos exámenes aportan valor a los usuarios ya que no se tienen que desplazar para examinarse y consiste en cuatro partes y tiene una duración aproximada de una hora.

Por último, cabe destacar que Duolingo cuenta con sistemas informáticos que van mejorando a medida que se introducen nuevas herramientas, lo que proporciona valor a la empresa a la hora de entrenar al algoritmo. Esto se traduce en una mejora en eficiencia para la elección más óptima entre todas las traducciones recibidas.

Por otro lado, reCAPTCHA se posiciona en el mercado como una herramienta de seguridad enfocada a la detección de *bots* y posibles actividades malignas. Se trata de una herramienta cada

vez más reconocida por los usuarios de las páginas webs y son presentados de cara al público como un paso adicional de seguridad. A su vez, esta herramienta está integrada en la página web de tal forma que el usuario no es redirigido a otra nueva página para realizar esta verificación, sino que esta se produce en la propia página web. Es decir, tiene una reputación reconocida en el mercado como herramienta de fácil uso contra actividades sospechosas.

En reCAPTCHA, la capacidad de respuesta del mercado se consigue gracias al análisis de los datos de la interacción de los usuarios con la herramienta. Es decir, el *Big Data* facilita en esta situación la observación de las respuestas que proporcionan los usuarios con respecto a los códigos generados. Esto permite analizar el formato idóneo dependiendo del tipo de cliente que tenga la página web. Es decir, algunos usuarios encontrarán más fácil la resolución del reCAPTCHA en formato imagen con respecto de la resolución de un código numérico.

En el caso de la fidelización, reCAPTCHA no establece un vínculo directo con los usuarios de las páginas web, ya que se trata de una herramienta de seguridad. En cambio, si funciona adecuadamente, los usuarios van a tener una visión positiva acerca de esta herramienta en cuanto a una experiencia más protegida. Esto a su vez es un elemento a favor de la empresa que implementa reCAPTCHA en sus sistemas, ya que puede dar confianza y seguridad a sus usuarios para seguir haciendo uso de la página web. A su vez, al igual que en el caso de Duolingo, se trata de un servicio cuyo cliente son las páginas web y por lo tanto a través del correcto funcionamiento del sistema, estos clientes van a mantenerse fieles al servicio al evitar tratar con *bots*. Podría decirse que se los clientes hacen uso de este sistema como método de prevención. Cabe destacar el segundo grupo de clientes con el que cuenta reCAPTCHA, que son las empresas que requieren de interpretación de caracteres o imágenes. Según Ahn (2008), se realizó una muestra de 50 artículos de *New York Times* que revelaron una eficiencia del servicio de un 99,1%, frente a la eficiencia de competidores de un 83,5%.

Ligado a esta última dirección se encuentra el aumento de ingresos que puede experimentar la empresa al dar una impresión a los usuarios de una página web segura y de fiar. Con el análisis de los datos la empresa puede adaptar el formato óptimo de reCAPTCHA para poder aportar una mayor comodidad de utilización a sus usuarios, y así facilitar su acceso y posible adquisición.

El descubrimiento de conocimientos se lleva a cabo por reCAPTCHA de tal forma que identifica actividades malignas o *bots* que no sería posible identificar sin la ayuda de esta herramienta. Es decir, reCAPTCHA saca a la luz información oculta para las empresas que cuenten con este sistema de seguridad impidiendo la utilización de los servicios con fines perjudiciales. A su vez, reCAPTCHA saca a la superficie la interpretación de caracteres o palabras que el reconocimiento óptico de caracteres ha sido incapaz de descifrar.

### 5.3. Conclusiones del caso y recomendaciones

Como se puede observar en el caso de Duolingo (**Tabla 2**), se identifica una nueva dirección para añadir en el marco de direcciones de valor. Esta nueva dirección gira en torno al concepto mencionado por Luis von Ahn de eficiencia social: *el beneficio social*. Esta dirección consiste en aportar valor a los clientes a través de los recursos con los que en este caso Duolingo cuenta: el *Big Data*. Es decir, sin contar con el apoyo de las empresas solicitantes de traducciones, Duolingo no sería capaz de ofrecer un servicio gratuito, y por consiguiente, aportar un beneficio social. En este caso, el beneficio social del *Big Data* está aportando valor a los usuarios al ofrecerles acceso gratuito y así contribuir a reducir la desigualdad social y económica.

Por tanto, Duolingo es una herramienta de aprendizaje de idiomas que ofrece a sus usuarios un servicio gratuito a cambio de las traducciones que realizan estos mismo. Podría considerarse que a los usuarios se les aporta datos no estructurados, que en este caso son traducciones de textos, que Duolingo luego reutiliza como fuente de traducción para vender traducciones profesionales a empresas. Es por ello, por lo que los usuarios de esta plataforma se consideran fuentes de valor para Duolingo, ya que serán estos los que estructuren estos datos. A su vez, Duolingo hace uso del *Big Data* para enriquecer su plataforma mediante el uso de técnicas que lo sitúan en su mercado objetivo. A través de las herramientas de *Big Data*, Duolingo pretende a su vez retener clientes, ya que al fin y al cabo son los creadores de las traducciones. Este aumento y retención de clientes le ha permitido obtener mayores ingresos y así invertir en la mejora de la infraestructura informática, lo que le permite seguir caminando hacia la consecución de uno de sus objetivos, que no es otro que traducir la web.

Aunque reCAPTCHA también haga negocio con la traducción de la web, cuenta con un enfoque diferente. Esta empresa por una parte proporciona servicios de protección a diferentes páginas web, que verifica que los usuarios son humanos y no robots. Con las respuestas obtenidas, reCAPTCHA contribuye a la traducción de artículos y libros. Nuevos conocimientos se generan cada vez que un usuario contesta un código, ya que pasan de ser caracteres no identificables a caracteres identificados. Es un modelo de negocio dual, por una parte, cobra por servicios de autenticación y por otra parte cobra por la traducción.

**Tabla 2: Direcciones de valor de Duolingo y reCAPTCHA**

	<b>Duolingo</b>	<b>reCAPTCHA</b>
Posicionamiento en el mercado	✓	✓
Capacidad de respuesta del mercado	✓	✓
Fidelización de la clientela	✓	✓
Desarrollo de competencias	✓	✗
Aumento de los ingresos	✓	✓
Aumento de la productividad	✗	✗
Ahorro de costes	✗	✗
Descubrimiento de conocimientos	✗	✓
Mejora de la infraestructura informática	✓	✗
<i>Beneficio social</i>	✓	✗

## 6. Conclusiones

### 6.1. Conclusión de objetivos

En cuanto a la tecnología, se puede concluir que está ya suficientemente madura para que sea empleada de forma masiva en el mundo empresarial, mejorando sustancialmente la productividad.

Esta tecnología permite la implantación de las herramientas de *Big Data* en las empresas, con las que se pueden analizar datos estructurados (que ya se analizaban en épocas anteriores al *Big Data*) y no estructurados que son los que aportan el valor diferencial a las estrategias empresariales.

En el caso de Duolingo y reCAPTCHA, el objetivo de traducir la *Web* es posible gracias a las aportaciones de los usuarios, a la tecnología y las herramientas de *Big Data* que están utilizando. Se crea un valor en ambas partes, ya que los usuarios pueden aprender idiomas o acceder a páginas Web de forma segura y por otra parte estas dos empresas se benefician económicamente al recibir las traducciones de los propios usuarios.

Se puede concluir que el *Big Data* es en estos momentos una de las fuentes de valor con mayor aportación a las estrategias empresariales.

## 6.2. Limitaciones

En cuanto a las limitaciones surgidas durante la realización de este trabajo, destaca que se trata de un tema novedoso y en constante desarrollo. El entorno empresarial y tecnológico en el que se desarrolla el valor estratégico del dato está en constante evolución. Las condiciones y factores externos pueden cambiar rápidamente, lo que puede afectar la aplicabilidad de los hallazgos y conclusiones en un futuro cercano.

Con respecto al caso real de Duolingo y reCAPTCHA, existen limitaciones en cuanto a la información pública de los procesos internos de ambas empresas. Esta información es de carácter privado, y, por lo tanto, para el estudio de ambos casos se ha hecho uso de información pública.

## 6.3. Futuras líneas de investigación

La conclusión personal que he sacado de la realización de este trabajo es que, aunque en un principio me pareció muy retador elegir un tema relacionado con el mundo de la Tecnología y el

*Big Data*, una vez hecho el trabajo se puede decir que es un tema muy interesante y que merece la pena seguir profundizando.

En especial una de las futuras líneas de investigación es el desarrollo del beneficio social como dirección en la creación de valor. Se trata de una dirección nueva que ayuda a las empresas a tener una motivación para proporcionar las mismas oportunidades a toda la sociedad.

Otra futura línea de investigación sería analizar como garantizar la Ética en la creación de valor del *Big Data*.

## 7. Referencias

- Ahn, L. V. (2012, Octubre 5). Entrevista con Luis von Ahn - creador de Duolingo y los Captchas (en español con subtítulos) . (B. Lewis, Interviewer)
- Ahn, L. v., Maurer, B., McMillen, C., Abraham, D., & Blum, M. (2008). reCAPTCHA: Human-based character recognition via web security measures. *Science*, 321(5895), 1465-1468.
- Arora, D., & Malik, P. (2015). Analytics: Key to Go from Generating Big Data to Deriving Business Value. *2015 IEEE first international conference on big data computing service and applications*, 446-452.
- Chen, D. Q., Preston, D. S., & Swink, M. (2015). How the Use of Big Data Analytics Affects Value Creation in Supply Chain Management. *Journal of Management Information Systems*, 32(4), 4-39.
- Chen, Y., Wang, W., Liu, Z., & Lin, X. (2009). Keyword Search on Structured and Semi-Structured Data. *Proceedings of the 2009 ACM SIGMOD International Conference on Management of data*, 1005-1010.
- Diamond, M., & Mattia, A. (2017). Data visualization: an exploratory study into the software tools used by businesses. *Journal of Instructional Pedagogies*, 18.
- Eberendu, A. C. (2016). Unstructured Data: an overview of the data of Big Data. *International Journal of Computer Trends and Technology (IJCTT)*, 38(1), 46-50.
- Elia, G., Polimeno, G., Solazzo, G., & Passiante, G. (2020). A multi-dimension framework for value creation through Big Data. *Industrial Marketing Management*, 90, 617-632.
- Erevelles, S., Fukawa, N., & Swayne, L. (2016). Big Data consumer analytics and the transformation of marketing. *Journal of Business Research*, 69(2), 897-904.
- Ermakova, T., Blume, J., Fabian, B., Fomenko, E., Berlin, M., & Hauswirth, M. (2021). Beyond the Hype: Why Do Data-Driven Projects Fail?. *Proceedings of the 54th Hawaii International Conference on System Sciences*, 5081-5090.
- Gavarri, S. L. (2016). El aprendizaje de lenguas extranjeras mediado por las TIC: aprender Inglés con Duolingo. *El toldo de Astier*, 7(12), 56-65.
- González-Padilla, P., & Galán Zori, C. (2021). Big Data Optimization in Zara: How Zara Will Optimize Its Process With Big Data Due to the COVID-19 Situation. *Advanced Digital Marketing Strategies in a Data-Driven Era*, 145-157.



- Grover, V., Chiang, R. H., Liang, T.-P., & Zhang, D. (2018). Creating strategic business value from big data analytics: A research framework. *Journal of management information systems*, 35(2), 388-423.
- Habib, H., Li, M., Young, E., & Cranor, L. (2022). “Okay, whatever”: An Evaluation of Cookie Consent Interfaces. *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, 1-27.
- Hacker, S. (2014). Duolingo: Learning a language while translating the web (Doctoral dissertation, PhD Thesis CMU-CS-14-116). *Carnegie Mellon University*.
- Lamba, H. S., & Dubey, S. K. (2015). Analysis of Requirements for Big Data Adoption to Maximize IT Business Value. *2015 4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions)*, 1-6.
- Liu, C.-h. (2015). A Conceptual Framework of Analytical CRM in Big Data Age. (IJACSA) *International Journal of Advanced Computer Science and Applications*, 1(6), 149-152.
- Müller, O., Fay, M., & Brocke, J. V. (2018). The effect of big data and analytics on firm performance: An econometric analysis considering industry characteristics. *Journal of Management Information Systems*, 35(2), 488-509.
- Marx, V. (2013). The big challenges of big data. *Nature*, 498(7453), 255-260.
- Merchant, R. M., Asch, D. A., Crutchley, P., Ungar, L. H., Guntuku, S. C., Eichstaedt, J. C., . . . Schwartz, H. A. (2019). Evaluating the predictability of medical conditions from social media posts. *Plos one*, 14(6).
- Munafó, F. (2019). La importancia de la gestión de datos y su impacto en el riesgo de crédito de instituciones financieras. *Revista de Investigación en Modelos Financieros*, 2, 25-38.
- Nam, T.-J., & Lee, W. (2003). Integrating Hardware and Software: Augmented Reality Based Prototyping Method for Digital Products. *CHI'03 extended abstracts on Human factors in computing systems*, 956-957.
- Petzold, B., Roggendorf, M., Rowshankish, K., & Sporleder, C. (2020). Designing data governance that delivers value. *McKinsey Digital*.
- Quy, V. K., Hau, N. V., Anh, D. V., Quy, N. M., Ban, N. T., Lanza, S., . . . Muzirafuti, A. (2022). IoT-Enabled Smart Agriculture: Architecture, Applications, and Challenges. *Applied sciences*, 12(7), 3396.

- Raguseo, E. (2018). Big data technologies: An empirical investigation on their adoption, benefits and risks for companies. *International Journal of Information Management*, 38(1), 187-195.
- Rajpurohit, A. (2013). Big data for business managers—Bridging the gap between potential and value. *IEEE International Conference on Big Data*, 29-31.
- Regulation, P. (2018). General data protection regulation. *Intouch*, 25, 1-5.
- Shamsuzzoha, A., & Raappana, H. (2021). Perspectives of business process ethics in data-driven marketing management. *Security and Privacy*, 4(6).
- Stobierski, T. (2021, February 2). 8 steps in the data life cycle. Retrieved from Harvard Business School Online: <https://online.hbs.edu/blog/post/data-life-cycle>
- Tanwar, M., Reena, D., & Khatri, S. K. (2015). Unravelling Unstructured Data: A Wealth of Information in Big Data. *2015 4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO)(Trends and Future Directions)*, 1-6.
- Tsai, C.-P., Chang, C.-W., Hsiao, H.-C., & Shen, H. (2022). The Time Machine in Columnar NoSQL Databases: The Case of Apache HBase. *Future Internet*, 14(3), 92.
- Urbinati, A., Bogers, M., Chiesa, V., & Frattini, F. (2019). Creating and capturing value from Big Data: A multiple-case study analysis of provider companies. *Technovation*, 84, 21-36.
- Wheeler, G. (2016). Machine Epistemology and Big Data. *The Routledge Companion to Philosophy of Social Science*, 341-349.
- Wu, X., Liang, L., & Chen, S. (2022). How big data alters value creation: Through the lens of big data competency. *Management Decision*, 60(3), 707-734.
- Yoo, Y., Henfridsson, O., & Lyytinen, K. (2010). Research Commentary—The New Organizing Logic of Digital Innovation: An Agenda for Information Systems Research. *Information Systems Research*, 21(4), 724-735.
- Yu, Q., & Yuan, J. (2022). TransBigData: A Python package for transportation spatio-temporal big data processing, analysis and visualization. *The Journal of Open Source Software*, 7(71), 4021.
- Zhang, D., Yin, C., Zeng, J., Yuan, X., & Zhang, P. (2020). Combining structured and unstructured data for predictive models: a deep learning approach. *BMC medical informatics and decision making*, 20(1), 1-11.

- Benny Lewis. [Benny Lewis] (2012). Entrevista con Luis von Ahn - creador de Duolingo y los Captchas (en español con subtítulos). *Youtube*.  
<https://www.youtube.com/watch?v=nLSpEbLu-lY>
- Luis von Ahn. [TED-Ed Español] (2016). Cómo aprender una lengua y contribuir a la sociedad | Luis von Ahn. *Youtube*. <https://www.youtube.com/watch?v=lt7yZB92McM&t=15s>
- Hacker, S. B. H. (2014). Duolingo: Learning a language while translating the web (Doctoral dissertation, PhD Thesis CMU-CS-14-116, *Carnegie Mellon University*, 2014. 5)
- Real Academia Española. (s.f.). Tecnología. *En Diccionario de la lengua española. Recuperado el 7 de junio de 2023, de <https://dle.rae.es/tecnolog%C3%ADa>*
- Gill, S. S., Tuli, S., Xu, M., Singh, I., Singh, K. V., Lindsay, D., ... & Garraghan, P. (2019). Transformative effects of IoT, Blockchain and Artificial Intelligence on cloud computing: Evolution, vision, trends and open challenges. *Internet of Things*, 8, 100118.
- Hill, C. A., Biemer, P. P., Buskirk, T. D., Japac, L., Kirchner, A., Kolenikov, S., & Lyberg, L. E. (Eds.). (2020). Big Data meets survey science: A collection of innovative methods. *John Wiley & Sons*.