

AF1: Información y alfabetización digital

Introducción al big data

Digitalización aplicada al sector productivo.

Módulo formativo sobre competencias digitales transversales básicas.



Índice

1. INICIO	3
1.1 Introducción	3
2. BIG DATA.....	4
2.1 Big data	4
2.2 Antecedentes del big data.....	8
2.3 Ciencia de datos y big data.....	9
2.4 Importacia del big data	10
2.5 Tecnologías de big data : recogida de información, procesamiento y almacenamiento	11
2.6 Un proyecto big data	12
2.7 4 V del big data	12
3. CIERRE.....	13
3.1 Resumen.....	13
3.2 Referencias bibliográficas.....	14

1. INICIO

1.1 Introducción

Las empresas buscan obtener información de su clientela, del mercado o de la economía para poder dar respuesta a sus necesidades de forma eficaz y rápida. Una forma de lograrlo es llevar a cabo campañas multitudinarias de manera personalizada y efectiva. **Big data** es el canal para conseguir toda esa información tan valiosa.

El big data incluye **muchas tecnologías** que han abierto las puertas hacia una nueva forma de trabajar con enormes cantidades de datos.

Esta unidad está dedicada a cómo se obtienen y qué utilidades tienen las **cantidades masivas** de datos que la tecnología es capaz de extraer, así como su gestión.

¿Qué vas a aprender en esta unidad?

OBJETIVO GENERAL DE LA ACCIÓN FORMATIVA
<ul style="list-style-type: none">Identificar, localizar, obtener, almacenar, organizar y analizar información digital, datos y contenidos digitales, evaluando su finalidad y relevancia para las distintas tareas profesionales.

OBJETIVO ESPECÍFICO DE LA UNIDAD DIDÁCTICA
<ul style="list-style-type: none">Definir el big data, enumerar sus principios y reconocer su importancia en la sociedad actual.

CONTENIDOS DE LA UNIDAD DIDÁCTICA
<h4>INTRODUCCION AL BIG DATA</h4> <ul style="list-style-type: none">1 ¿Qué es el big data?2 Antecedentes del big data3 Ciencia de datos y big data4 Importancia del big data5 Tecnologías de big data: recogida de información, procesamiento y almacenamiento6 Un proyecto big data7 4 V del big data

2. BIG DATA

2.1 Big data

Big data, es un término que viene del inglés que se podría traducir al español como "datos masivos". Se trata de un concepto relativamente nuevo y se refiere al **gran conjunto de datos**, que pueden llegar a ocupar varios **gigabytes** o **terabytes**, y a su análisis por parte de las compañías u otras organizaciones para **obtener información relevante y poder tomar decisiones**.

Hay que considerar que el big data **no es una tecnología específica**, sino que engloba varias tecnologías complejas dedicadas principalmente a recopilar cantidades enormes de datos y procesarlos para extraer información útil.

Por ejemplo, los sensores que miden los niveles de dióxido de carbono (CO₂) en el aire recopilan sistemáticamente una serie de datos que dan información sobre los niveles de contaminación que se pueden llegar a alcanzar en las ciudades.

Big data es un conjunto de tecnologías que consisten en la recopilación de grandes cantidades de datos y su procesamiento para transformarlos en información, y de ahí en conocimiento para tomar decisiones acertadas y pasar a la acción.



El reto para las organizaciones está en saber manejar esa gran proporción de datos y ser capaces de extraer información útil a partir de ellos.

¿Qué permite el análisis de datos?

Lo que hace que el big data sea tan beneficioso es su **capacidad de dar respuestas** a la hora de tomar decisiones de cara a mejorar el negocio o cumplir con los objetivos de las organizaciones a partir de los datos que se manejan.

Un análisis de los datos con técnicas de big data permite:

- Identificar **nuevas oportunidades** de negocio.
- **Encontrar patrones** repetitivos y predecir cambios.
- Implementar **procesos automatizados** de producción para reducir costes.
- Satisfacer **necesidades individuales** y personalizadas de la clientela.
- **Identificar tendencias de gustos**, intereses y de consumo de la población.
- **Tomar decisiones** que contribuyan a mejorar la calidad de vida de las personas.

Siguiendo con el ejemplo anterior, el análisis de los datos sobre CO2 nos da información de los niveles de contaminación, lo que puede servir de base para que los ayuntamientos tomen medidas para mejorar la calidad del aire en las ciudades, como peatonalización de calles o creación de carriles bici.

¿Dónde se aplica el big data?

Las técnicas de big data se aplican principalmente en áreas o servicios que recopilan gran cantidad de datos como por ejemplo los gestores de correo electrónico, los navegadores y buscadores web, los servicios de **mensajería SMS** o las **redes móviles**.

También en aplicaciones que detectan temperaturas, niveles de contaminación, pasos que damos al día o nuestra ubicación, como las que instalamos en nuestros smartphones.

Como puedes observar, el big data está muy presente en muchas de las tecnologías y aplicaciones que utilizas a diario.

Un ejemplo de big data es el proyecto LUCA de Telefónica que analiza la información que recogen las antenas de telefonía en las carreteras para reducir los atascos o descubrir los destinos turísticos favoritos de la ciudadanía.

El gigante de las redes sociales: Facebook + WhatsApp + Instagram

Para que te hagas una idea de la importancia del big data, Facebook, también dueño de Messenger, WhatsApp e Instagram, ha basado su éxito como modelo de negocio en la **recopilación y análisis de toda la información** que recogen estas aplicaciones, con el consentimiento de las personas usuarias, para llevar a cabo campañas personalizadas que les han reportado grandes beneficios económicos.

La sencilla opción "Me gusta" que popularizó Facebook supuso un hito a la hora de recopilar información muy valiosa sobre los gustos y preferencias de las personas.

Otro ejemplo lo podemos encontrar en la red social de empleo LinkedIn en el que cada persona registrada recibe ofertas personalizadas de trabajo de acuerdo a la información de su perfil, relaciones que establece con sus contactos y publicaciones que realiza.

¿Cómo recopilan la información las aplicaciones y sitios web?

1. Cesión de datos personales

En la unidad didáctica 1, veíamos que al dar de alta una cuenta de correo electrónico la aplicación pedía cierta información personal a la persona usuaria a cambio de poder utilizar el servicio y recibir recomendaciones o información personalizada. Estos datos que facilitamos son los que emplean muchas aplicaciones web en sus técnicas de big data.

Como persona usuaria debes saber que tienes el derecho de recibir avisos sobre los criterios de privacidad y que tienes el derecho de gestionarlos como mejor te convenga, incluso de cancelar tu consentimiento de ceder tus datos, si así lo deseas.

En el caso de que detectes un uso indebido de tus datos personales puedes denunciar a la compañía por incumplimiento de la Ley de Protección de Datos.

Reflexiona...

La información personal tiene mucho valor para las compañías porque dice mucho de ti y de tus preferencias, pero se trata de información privada y solo tú decides cómo cederla o compartirla. Revisa siempre las políticas de privacidad y uso de datos personales de las aplicaciones y servicios web.

2. Cookies de seguimiento

Nuestro comportamiento al interactuar con las opciones que ofrecen las aplicaciones y sitios web también da mucha información personal. Para recopilar toda esta información están las cookies de seguimiento.

Cuando accedes a cualquier página web o aplicación móvil, aparece una petición de consentimiento del uso de cookies. Una cookie es un paquete de datos que el navegador web almacena en tu dispositivo cuando visitas una página web. Estos son algunos ejemplos:



- **Ubicación**

Al aceptar las cookies de navegación, proporcionas la información del lugar físico en el que te encuentras.

- **Información de las búsquedas**

Las cookies obtienen datos sobre la información que buscas, las palabras clave que usas, los resultados que visitas, las tiendas en las que compras, etc.

- **Los "me gusta"**

Muchas aplicaciones, sobre todo, las redes sociales, pero también publicaciones digitales como periódicos o revistas, pone a disposición de las personas usuarias la opción de "Me gusta" (like en inglés) lo que les permite recopilar datos a través de las cookies sobre gustos, intereses y preferencias.

- **Las páginas que se visitan**

Las cookies almacenan y rastrean todas las páginas que visitas: cuándo lo haces, durante cuánto tiempo e incluso si te instalas una aplicación o haces una compra.

Reflexiona...

La Ley de Protección de Datos Personales y Garantía de los derechos digitales, obliga a todas las webs a notificar a la persona usuaria su política de recopilación y uso de la información que lleva a cabo del tal modo que la persona interesada pueda aceptarla o rechazarla.

3. Identificación de contactos cercanos

Has de ser muy consciente de que cuando etiquetas a alguien con **nombre y apellidos** en una foto que compartes en redes sociales, estás identificando a la persona y ofreciendo información a la aplicación sobre quiénes son tus vínculos más cercanos.

Por otro lado, las técnicas de **reconocimiento facial** están cada vez más avanzadas por lo que el reconocimiento de las personas a través de fotografías será también posible.

2.2 Antecedentes del big data.

El matemático Alan Turing es considerado el padre de la computación moderna y fue pionero en desarrollar técnicas y procesos relacionados con la creación de conocimiento a partir de la analítica de datos masivos.

Algunos antecedentes del estudio de datos masivos los podemos situar en la figura de **Alan Turing** (Reino Unido, 1912 - 1954) que ya teorizaba en los años 30 sobre la posibilidad de crear una máquina capaz de realizar **tareas automatizadas** a partir de la recolección de datos e identificación de patrones.

El científico lideró un equipo de investigación durante la Segunda Guerra Mundial que logró descifrar parte de los mensajes codificados que enviaban los alemanes a través de **Enigma, una máquina especial para el cifrado de comunicados**, lo que permitió acortar la guerra entre 2 y 4 años.

En 1948 se publicó la **Teoría Matemática de la Comunicación**, de Claude Shannon, en la cual se determinaban los requisitos mínimos para transmitir información a través de canales. Este trabajo contribuyó a minimizar el volumen de datos. Aunado a esto, años después se desarrolló el término de "memoria virtual", que tiene que ver con el almacenamiento ilimitado e infinito de los datos: el procesamiento de los datos se podía realizar sin restricciones en materia de hardware ("máquina", por así decirlo).

La primera acción con datos masivos se produjo en el **censo en los Estados Unidos en los años 80**. Tardaron **ocho años solo en organizarlos y tabularlos**. No obstante, tiempo después calcularon que, con la tecnología existente y la importancia del trabajo humano y manual en esa tarea, esta labor hubiera necesitado 10 años para ser terminada. Por ello, la búsqueda de adelantos metodológicos y tecnológicos era fundamental.

Tiempo después, Herman Hollerith inventó la **máquina tabuladora**. Con ella, organizar el censo de los EE.UU. se reducía en el tiempo, pues podría tenerse en el lapso de un año. Sin embargo, así como este proceso avanzaba, también lo hacían los datos. Esto hacía que incluso la máquina tabuladora tuviera inconvenientes.

¿Sabías qué?

La máquina tabuladora tenía como objetivo "contar" los datos del censo en EEUU. Para ello, utilizaba un sistema, un contador, un clasificador y un tabulador.

"Los datos son el petróleo del siglo XXI. El despliegue de sensores y el incremento de la capacidad del procesamiento son claves en la transformación de muchos sectores y en la creación de un mundo más medible y programable".
César Alierta, presidente Fundación Telefónica

2.3 Ciencia de datos y big data

La ciencia de datos es la disciplina que convierte los datos en conocimiento útil.

Ya hemos visto que el objetivo del big data es **transformar los datos en información útil**. Para ello el big data se basa en la **ciencia de datos** que combina diferentes técnicas de las matemáticas y ciencias informáticas con el fin de unificar estadísticas y datos de distinta naturaleza para interpretarlos correctamente.

La digitalización produce tal cantidad de información, que solo es posible analizarla mediante la ciencia de datos que utiliza programas y procesos específicos para ello.

Y entonces, ¿qué es un dato?

Un dato es la unidad mínima de información. Un dato se define como la representación de una variable (cualitativa o cuantitativa). En otras palabras, un dato es información.

Origen de los datos	Ejemplos de datos
Web y redes sociales	Cada "me gusta" (<i>like</i>), cada comentario, cada vez que compartes un artículo o guardas una foto, cada vez que ves una publicación y no interactúas con ella, cada vez que ves una publicación y sí interactúas con ella... Básicamente, todo lo que haces en la red y en las redes sociales genera datos. Lo mismo ocurre con las páginas web: el tiempo que estás en un sitio web, las páginas que visitas, las compras de comercio electrónico o las búsquedas que realizas en Google y otros buscadores. Todo ello genera datos.
Personales	1. Biométricos: Desde los más sencillos como sexo, altura, peso, a datos como el reconocimiento facial. 2. Generados por cada persona: Cuando recibes una llamada comercial en la que se te pide permiso para grabar la conversación, cuando respondes a un formulario electrónico, etc.
M2M	Cuando dos máquinas comparten información, por ejemplo: los electrodomésticos inteligentes.
Big Transaction Data	Por ejemplo, registros de facturación de una empresa, inventarios, carteras de clientes. Son datos estructurados y no estructurados .

Tipos de datos

Actualmente, se pueden distinguir dos tipos de datos: estructurados y no estructurados.

- **Datos estructurados.**

Datos estructurados son aquellos que tienen una **estructura y una organización** fácil de identificar. Suelen ser archivos de texto que se almacenan en formato de tablas, hojas de cálculo o **bases de datos relacionales** con títulos reconocibles para cada unidad de información.

Por ejemplo: un informe de ventas del último mes.

- **Datos no estructurados.**

Este tipo de datos, como su nombre indica, no tienen una estructura identifiable y se almacenan en múltiples formatos de imagen, correos electrónicos, mensajes de texto o información multimedia como audios o videos.

Dentro de esta categoría se incluye la información generada por máquinas como las imágenes de satélite o la información que generan los sensores de tráfico o meteorológicos, por ejemplo.

2.4 Importancia del big data

Los datos que se manejan para esta tecnología son realmente abrumadores: hablamos de miles de millones de móviles y dispositivos conectados y terabytes de información transmitida alrededor de todo el mundo en tiempo real.

En este contexto, no solo las empresas se están dando cuenta de la necesidad de regirse por los datos para lograr **beneficios en sus negocios**, también las organizaciones e instituciones públicas y gubernamentales empiezan a mirar al big data como una herramienta que puede contribuir a **lograr los objetivos marcados para mejorar la vida de las personas**.

Los datos son indispensables para todo en la actualidad, y se están convirtiendo en una pieza clave para la toma de decisiones estratégicas.

El big data y las empresas

No cabe duda de que el big data es una herramienta de gran utilidad para las empresas pues el análisis de los datos que, por ejemplo, proporcionan su clientela sobre sus productos o servicios permite a las compañías adaptarlos a sus intereses específicos o identificar nuevas necesidades a las que dar respuesta.

Al incorporar big data en las empresas es posible lograr nuevos objetivos, nuevas líneas de negocio y capacidades:

1. **Costes:** se detectan e implementan procesos más eficientes que reducen costes.
2. **Toma de decisiones eficaz y rápida:** el análisis de los datos permite determinar, incluso predecir, causas de fallos, problemas o defectos en la producción.
3. **Nuevos productos y servicios:** permite conocer las necesidades concretas de la clientela y su nivel de satisfacción, pudiendo ofrecer productos y los servicios adaptados a sus necesidades.

Cuando buscas en la red, por ejemplo, alojamientos y planes para un viaje, no es extraño que después te aparezcan anuncios en páginas web y redes sociales asociados a ese viaje que has estado buscando, como anuncios de trajes de baño, artículos para la playa o publicidad de planes de viajes. Esto es el resultado del uso del big data.

El big data y las organizaciones institucionales

El big data también es una herramienta de gran utilidad a la hora de identificar y medir problemáticas para poder ofrecer soluciones alineadas con los Objetivos de Desarrollo Sostenible (ODS) de la Agenda 2030.

Por ejemplo, los datos de satélites ayudan a identificar zonas de sequía y se ayuda antes a las poblaciones afectadas. También los datos de sensores en las fuentes de agua informan sobre si es potable. El uso de teléfonos móviles y las redes sociales avisan de situaciones de desastres (terremotos, inundaciones) y permiten enviar las ayudas más rápida y directamente a los lugares donde se necesitan

2.5 Tecnologías de big data : recogida de información, procesamiento y almacenamiento

El ordenador Q System One de IBM ha logrado resolver en unos tres minutos un problema de generación de números aleatorios en el que los superordenadores clásicos más potentes del planeta habrían invertido 600 millones de años. Revista Science. Diciembre 2020.

¿Cómo se analizan e interpretan esos datos?

Al inicio de esta unidad didáctica se explica que el big data aglutina varias tecnologías que se podrían resumir en las siguientes:

1. **Tecnologías para recopilar datos** relacionados con las visitas a páginas web, publicaciones en redes sociales y registros de llamadas.
2. **Tecnologías orientadas al procesamiento de la información** que se encargan de preparar y etiquetar los datos para que puedan ser analizados.

3. **Tecnologías de almacenamiento de datos** para salvaguardar toda esa información y que sea accesible.

"La tecnología es importante, pero lo único que realmente importa es qué hacemos con ella". Muhammad Yunus, premio Nobel de la Paz 2006 por el desarrollo del "Banco del Pueblo" dirigido a personas pobres y mujeres de la India.

2.6 Un proyecto big data

Fases de un proyecto big data

Todo proyecto de big data, sea cual sea su aplicación o caso de uso, pasa por cuatro fases principales que a continuación pasamos a describir brevemente.

1. Definición de los recursos básicos: objetivo y equipo

El primer paso será determinar la finalidad o los objetivos que se desean lograr aplicando las técnicas de big data. Así, será más sencillo decidir los recursos que se necesitarán y desglosar un plan de tareas.

2. Determinar los datos a integrar

En esta etapa del plan se deciden las fuentes que van a proporcionar los datos que se van a analizar, es decir, de dónde se van a extraer los datos.

3. Integrar los datos

En este punto se define la tecnología big data que se va a implementar para el análisis e integración de los datos obtenidos.

4. Estudio de datos dentro del plan

La fase final del proyecto es el análisis de los propios datos para extraer la información sobre la que se basará las futuras tomas de decisiones.

Aplicaciones como Cabify o Uber utilizan el dato de la ubicación del automóvil para informar a la persona usuaria el tiempo que va a tardar en llegar. Traducir ese dato en información útil para las personas usuarias es posible gracias al uso de big data.

2.7 4 V del big data

Todas las acciones que realizamos a través de las redes sociales, aplicaciones web, correos electrónicos, etc. generan datos. Cada vez que haces una transacción bancaria, una compra en una tienda online o buscas algo en Internet se queda registrado.

Para dar **sentido a todos estos datos**, es necesario comprender su composición, por lo que los analistas de datos se apoyan en el **método de las 4 V del Big Data** para fijar

las **propiedades** de estas gigantescas bases de datos y poder asegurarse de que de ellas podrán obtener conclusiones válidas.

- **Volumen**

Además de los datos internos propios de la empresa, también se usan datos externos como la web, redes sociales, interacciones con la clientela o incluso datos procedentes de procesos automatizados que realizan máquinas en algunas compañías.

- **Variedad**

Los datos poseen gran cantidad de formatos, como imágenes, vídeos, metadatos, etc. y provienen de diferentes fuentes, como webs, redes sociales, bases de datos, correos electrónicos, etc. Contar con datos muy diversos permitirá cubrir un amplio rango de casuísticas.

- **Velocidad**

El proceso que va desde la recopilación de datos hasta el retorno del análisis debe ser ágil y en tiempo real.

- **Veracidad**

Distinguir los datos válidos para obtener información útil para lograr objetivos de lo que es ruido.

¿Sabías que...?

Justin King era consejero delegado de la cadena de supermercados británica Sainsbury, y explicaba cómo el manejo de los datos establecía relaciones curiosas. Por ejemplo, mostraba una relación entre las compras de productos dietéticos y que la persona estuviera planeando irse de vacaciones. Y esa información sirvió a la cadena de supermercados, por ejemplo, para hacerle llegar publicidad directa de bronceadores.

3. CIERRE

3.1 Resumen

¿Qué les ha ocurrido a nuestros protagonistas?

Aunque el negocio de Pedro y Javier es pequeño y no genera una masa de datos tan grande como para aplicar técnicas de big data, nuestros protagonistas han comprendido en qué consiste esta tecnología y lo importante que es en la sociedad actual. Además, han adquirido conciencia de las implicaciones y efectos que producen nuestras acciones en el mundo digital.

¿Y qué has aprendido a hacer?



Pedro y Javier se han introducido en las tecnologías del big data y están tomando una nueva conciencia de lo que es la información y de cómo se puede generar conocimiento en un mundo digitalizado. Pero todavía podemos dar un paso más, ya que el big data y su capacidad de reconocer patrones y hacer predicciones nos dirige directamente hacia un territorio apasionante: la inteligencia artificial. No es ciencia ficción, ¿quieres saber en qué consiste? En la siguiente Unidad Didáctica te lo contamos.

3.2 Referencias bibliográficas

A continuación, puedes ver la relación de recursos (artículos, estudios, investigaciones, páginas web...) que se han consultado y citado para elaborar el contenido de esta Unidad Didáctica:

- Agencia Estatal Boletín Oficial del Estado. (2018). Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales. Recuperado de: <<https://www.boe.es/buscar/act.php?id=BOE-A-2018-16673>> [20/01/2022].
- Raconteur. A day in data. 2017. Recuperado de <<https://www.raconteur.net/infographics/a-day-in-data/>> [20/01/2022].
- Schwab, K., Foro Económico Mundial, 2017. La cuarta revolución industrial (2.a ed.). Debate.

- Vodafone. Big data: desde los inicios hasta hoy, 2018. Recuperado de: <<https://ideasparatuempresa.vodafone.es/big-data-desde-los-inicios-hoy/>> [20/01/2022].