



Aproximación al Big Data. Análisis de su posible utilización en la universidad pública

Antonio Sánchez-Luengo Leví

Curso 2017-2018

Directora: Dra. Soledad María Martínez María-Dolores

Trabajo Fin de Estudios para la obtención del título de Graduado en Administración y Dirección de Empresas

Índice:

1. Introducción	3
2. Big Data	4
2.1. ¿Qué es Big Data?	4
2.2. Historia	6
2.3. El Big Data y la empresa en la actualidad	7
2.4. Datos	9
2.4.1. Estructura	9
2.4.2. Origen	11
2.4.3. Almacenamiento	15
3. Casos de éxito en la aplicación de Big Data	17
3.1. Utilización del Big Data en la educación internacional	20
4. Análisis de la situación de las universidades españolas	23
4.1. Universidad Pública en España	23
4.2. Universidad privada	25
4.3. Universidad pública contra universidad privada	26
5. Posible implantación en la UPCT	28
5.1 Situación y herramientas actuales	28
5.2 Legislación vigente	30
5.2.1. Ley de Protección de Datos	30
5.2.2. Reglamento General de Protección de Datos	31
5.3 ¿Qué aportaría y qué supondría el uso del Big Data en la UPCT?	32
6. Conclusiones	36
7. Bibliografía	38

Índice de figuras:

Figura 1: Exabytes generados mensualmente	5
Figura 2: Ejemplos según origen de los datos	11
Figura 3: El avance de internet en el mundo	12
Figura 4: Aplicaciones en las que pasamos el tiempo	13
Figura 5: ¿Qué sucede en internet en 60 segundos?	15
Figura 6: Ejemplo de formato documental	16
Figura 7: Ejemplo de formato grafo	17
Figura 8: Recomendaciones Netflix	18
Figura 9: Recomendaciones Amazon	19
Figura 10: Número de universidades por CCAA	24
Figura 11: Estudiantes matriculados curso 2017-2018	25
Figura 12: Grados impartidos curso 2017-2018	26
Figura 13: Número de universidades privadas por CCAA	28
Figura 14: Tareas realizadas por OPADA 2018	30

1. Introducción.

Este Trabajo Fin de Estudios (TFE) se enmarca dentro de los estudios de Grado en Administración y Dirección de Empresas, impartidos por la Universidad Politécnica de Cartagena (UPCT).

Se ha elegido este tema para realizar dicho TFE por la gran importancia que tiene la diferenciación para crear un valor añadido en las empresas, ya sean públicas o privadas. Además, como alumno que ha participado activamente durante sus años en la universidad en propuestas como la formación de la Asociación Juvenil de Estudiantes Emprendedores de Cartagena (AJEEC) y en organismos estudiantiles tales como la delegación de estudiantes, sabemos que resulta determinante la importancia que tiene para la universidad proporcionar al estudiante una calidad educativa cada vez mayor, mejorando los procesos de aprendizaje y su experiencia durante su etapa de educación superior.

Así pues, con el presente TFE se pretende explorar, si es posible, la utilización, a modo de experiencia piloto por parte de Universidad Politécnica de Cartagena, del Big Data con diferentes fines como podrían ser la captación de nuevos alumnos o la retención de los existentes. Así mismo podría ser de gran utilidad la creación de otros estudios específicos más demandados tales como especialidades dentro del programa de estudios vigente.

Para ello, este trabajo desarrolla y analiza tres puntos clave que llevará a conclusiones que sirven de argumento para proponer el uso de este tipo de herramientas en la UPCT.

En primer lugar, se analiza en profundidad qué es el Big Data, cómo se estructura, cómo se obtienen los datos y la forma en que éstos se presentan, así como la importancia que tienen.

A continuación, se dará una visión sobre el Sistema Universitario Español y se analizará el Big Data dentro de las universidades, tanto públicas como privadas, exponiendo casos de trabajo con esta herramienta para mejorar la experiencia tanto para el estudiante como para la propia universidad.

Para dar una visión más amplia sobre el tema, seguidamente se expondrán algunos casos de éxito de organizaciones tanto públicas como privadas, sobre los que se fundamentará el estudio.

Por último, el trabajo persigue señalar si el uso del Big Data para la UPCT como conclusión nuestra podría significar una diferenciación sustancial, que lograría un avance muy amplio tanto en la calidad de educativa para los estudiantes como en la mejora de la reputación y la visibilidad de esta entidad educativa

2. Big Data.

En este primer apartado nos introduciremos en la tecnología Big Data: en qué consiste, cómo funciona, de qué se compone y cómo se genera. Para ello empezaremos viendo qué entendemos por Big Data en la actualidad y cómo ha evolucionado este concepto. A continuación, hablaremos de forma somera de la historia de esta herramienta: cómo surgió, qué implica y cuáles son sus principales características. Posteriormente, profundizaremos en lo realmente importante en el Big Data, los datos: cómo se encuentran, qué tratamientos se les da y su utilidad. Para terminar, hablaremos del almacenamiento, explicando los distintos tipos que existen y para qué se utiliza cada uno de ellos.

2.1. ¿Qué es Big Data?

Existen multitud de definiciones para exponer qué es y en qué se basa el Big Data. Según McKinsey Global Institute (2011) “el Big Data se refiere al conjunto de datos cuyo tamaño está más allá de las capacidades de las herramientas típicas de software de bases de datos para capturar, almacenar, gestionar y analizar”. También se podría decir que el Big Data es la confluencia de una multitud de tendencias tecnológicas que han ido surgiendo desde la primera década del siglo XXI hasta los años 2011-2012 que ha sido cuando ha irrumpido en organizaciones y empresas (Joyanes, L. 2013).

Según Laney (2010) esta inmensa cantidad de datos tienen tres características en común, conocidas como “Las 3V del Big Data”: variedad, volumen y velocidad. A estas tres características, IBM ha añadido una cuarta “V” que es la veracidad y con el paso de los años y el mayor conocimiento de Big Data, se ha añadido una V más, que se refiere al valor.

Vamos a ver cada una de ellas con más detenimiento:

Volumen: hacemos referencia al volumen por la gran cantidad de datos que se generan diariamente en el mundo. Desde hace años las grandes empresas tecnológicas manejan una gran cantidad de datos. Todos ellos son almacenados, pero sólo una pequeña parte de ellos son analizados por las empresas o instituciones con el fin de obtener un posicionamiento mejor respecto a sus competidores o con el simple objetivo de mejorar el servicio prestado. Se espera que en el año 2020 se alcancen los 35 Zettabits (ZB) almacenados en el mundo (Esmartcity, 2017). Este dato nos da una idea de la importancia que tiene esta información para las organizaciones en un plazo muy pequeño de años.

En el siguiente gráfico vemos la evolución del tráfico de datos por mes en el mundo:

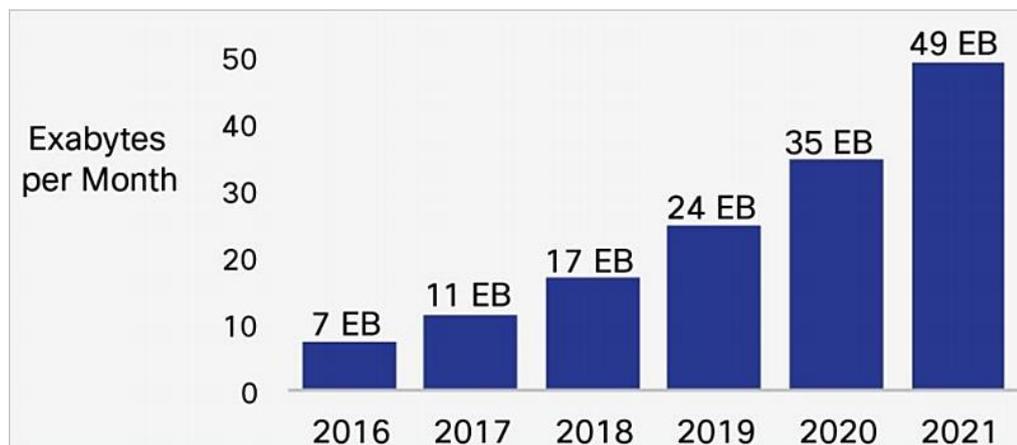


Figura 1: Exabytes generados por mes. Fuente: Esmartcity.es (2017)

Variedad: cuando hablamos de variedad nos referimos a que los datos utilizados en el Big Data no provienen de una sola fuente ni encontramos un solo tipo de datos. El Big Data utiliza información que proviene desde la navegación por internet hasta de las imágenes grabadas por las cámaras de tráfico en tiempo real. Todos estos datos que recopilamos no tienen una estructura fija, es decir son datos sin organización alguna, que necesariamente para su análisis han de poder ser estructurados con modelos y aplicaciones destinadas a ello. Por otro lado, hablaremos de datos estructurados cuando estos presenten un esquema fijo.

La variedad que presentan puede suponer un inconveniente ya que al recopilar la información de tantas fuentes distintas y la complejidad que presentan los datos

puede crear un conflicto con la seguridad y la vulnerabilidad de la privacidad. Es por esto por lo que hablaremos más adelante sobre la legislación actual que compete al Big Data, tanto a nivel estatal como comunitario.

Velocidad: No nos referimos con esto a la velocidad a la que son almacenados los datos, si no a la velocidad a la que se generan. Como comentábamos antes, cada día se generan más datos y cada vez a mayor velocidad, Como dijo Harriet Green (directora de la división de “Internet de las cosas” de IBM) “en los dos últimos años la sociedad ha generado más datos que en toda la historia”. Esto hace que para las organizaciones sea imprescindible adaptarse para lograr su análisis en tiempo real.

Veracidad: esta característica es la que como comentábamos anteriormente añadía IBM y hace referencia a la necesidad que tienen las organizaciones de que los datos obtenidos y analizados sean ciertos para que puedan ayudarles a tomar las decisiones más adecuadas. La veracidad de los datos supone un reto para el Big Data planteado para los próximos años debido al aumento del volumen y la variedad de los datos.

Valor: Muchos expertos han hecho hincapié en esta característica. Se tiene la idea equivocada que cuantos más datos, más valor se tiene y eso no es realmente así. El valor no lo produce la cantidad sino la posibilidad de análisis y las decisiones que puedan ayudar a tomar esos datos. Cuanto mayor análisis permitan los datos obtenidos, mayor y mejor toma de decisiones tendrá la empresa, y por tanto mayor será su valor (Instituto de Ingeniería del Conocimiento, 2016).

2.2. Historia.

En 1997 fue usado por primera vez el término “Big Data”. Aparece en un artículo publicado por los investigadores de la NASA Michael Cox y David Ellsworth, y es en ese momento cuando da comienzo la gran revolución del análisis de los datos. Pero fue con anterioridad cuando los investigadores se dieron cuenta de que el análisis de los datos y el almacenamiento de estos podía producir beneficios a las entidades que los manejasen.

En la década de los 40, investigadores estadounidenses advertían sobre la gran cantidad de datos que se generaban año tras año en el mundo. En la década de los 50 algunos especialistas comenzaron a hablar de la explotación de estos datos como “una necesidad de analizar con el objetivo de tomar decisiones inmediatas para tener un mejor futuro y crear *santuarios* de datos que contaran la historia más allá de las bibliotecas” (Paula Lara. 2017).

En 1999 Peter Lyman y Hal R. de la Universidad de California en Berkeley deciden publicar un estudio en el que se indicaba la cantidad de información nueva y original creada en el mundo en un sólo año, siendo 1,5 exabytes.

En 2001, Doug Laney menciona las 3V, que como hemos visto en el apartado anterior y que siguen vigentes. Pero no fue entonces cuando se produjo la revolución del Big Data, sino en el año 2006 cuando el buscador Google tuvo la necesidad de indexar¹ la web a los niveles que exigía el mercado. Para poder llevar a cabo estas acciones llegó Hadoop, un gran número de pequeños ordenadores, cada uno de los cuales se encarga de procesar una porción de información; posteriormente fue cedido con licencia Open Source² a la Apache Software Foundation, quien lo distribuyó libremente e hizo que se produjera la citada revolución en el análisis de datos.

Esta revolución continua en la actualidad gracias a las mejoras en la tecnología (cada vez tenemos sensores y gadgets más pequeños que generan a su vez más datos, como pueden ser los Smart Watch, Smart Phone, Smart band, etc.). y al auge de las redes sociales y blogs, que generan cada día una cantidad inmensa de datos.

2.3. El Big Data y la empresa en la actualidad.

Las empresas se están empezando a dar cuenta de la necesidad de utilizar el Big Data como elemento determinante en su toma de decisiones para mejorar sus resultados. Actualmente, son las grandes empresas tecnológicas como Facebook o Google las que más uso hacen de los datos para llevar a cabo sus políticas. Esto no es casualidad, ellas mismas son las principales generadoras de datos diarios en el mundo. Con cada búsqueda en Google o cada “me gusta” en Facebook estamos proporcionando una información muy valiosa sobre nosotros, nuestros gustos y

¹ Según la [RAE](#), indexar significa “registrar ordenadamente datos e informaciones, para elaborar su índice”.

² Software desarrollado y distribuido libremente y con acceso al código fuente programado.

aficiones, por dónde nos movemos, con qué asiduidad o incluso qué comemos y dónde lo hacemos. Pero no se quedan ahí, para poder ofrecernos más cosas y que se adecúen a nosotros de mejor forma, aparte de utilizar los datos que nosotros le proporcionamos también compran datos a otras compañías. Esto último sobre todo lo hacen las empresas que utilizan el Big Data pero que no obtienen los suficientes datos propios como para obtener una imagen lo suficientemente real para tomar decisiones correctas.

Aunque cada vez el Big Data está más integrado en la empresa y su entorno, muchos directivos creen que su uso no está creando un impacto económico real. Según comenta Carme Artigas, CEO de Synergic Partners en una conferencia durante la Big Data Week en Madrid (2017): “El reto de las empresas es adoptar Big Data a nivel corporativo y en procesos de negocio para lograr un impacto económico real”. Según esta autora esto sucede porque no es fácil para la empresa encontrar talento analítico que tenga la capacidad de realizar todas las acciones necesarias para el Big Data y “cuando la tiene, por lo general se concentran en un departamento en particular, de manera que el talento no se distribuye por igual en toda la organización”.

Para poder llevar a cabo lo expuesto por Carme Artigas, muchas empresas deben realizar un esfuerzo económico enorme que posiblemente no pueda asumir. A modo de solución se han lanzado una serie de soluciones para empresas y para la Administración Pública de la mano de Vodafone y CARTO (empresa creadora de una plataforma para convertir los datos de ubicación en resultados comerciales con el mismo nombre). A través de este acuerdo, se le proporciona a las empresas y las Administraciones Públicas información sencilla y visual haciendo que no sea necesario contratar a un experto en Big Data y toda la tecnología necesaria para el análisis ahorrando así mucho dinero a la empresa.

Hoy en día todos tenemos un móvil con el que vamos a todos sitios, llamamos, mandamos mensajes, mails y también pasamos tiempo de ocio en las redes sociales o con juegos. Todas estas acciones dejan información sobre nosotros, es decir, genera “datos” que como hemos afirmado ya cualquier empresa que los obtenga puede utilizar en su beneficio.

2.4. Datos.

Cuando hablamos de Big Data lo realmente importante son los datos, de todo tipo, con todos los formatos posibles y con cualquier estructura. En Big Data cualquier dato es válido, aunque proporcionan una cantidad y una calidad de información diferente unos de otros. Como vimos en puntos anteriores, la cantidad diaria generada crece cada año estimándose la creación de 774,2 Petabytes diarios para el año 2019 (Esmartcity, 2017). Que la cantidad de datos crezca cada día les proporciona una importancia cada vez mayor, que las organizaciones tienen que empezar a saber gestionar con el fin de crear características diferenciadoras, con el fin de aplicarlos en nuevas estrategias empresariales, más adaptadas al siglo XXI. La manera de obtenerlo también resulta diferente en la actualidad, si bien antes se conseguían a través de lo que se generaba en la web o algunos sensores aislados, ahora casi todo puede generar datos, desde una estación meteorológica, pasando por un reloj inteligente, hasta un sensor de medio de paso de vehículos en una carretera. Los datos se pueden clasificar de diversas maneras, a fin de simplificar su entendimiento, vamos a diferenciarlos según su estructura y según su origen.

2.4.1. Estructura.

En alguna definición se habla de datos estructurados, semi estructurados y no estructurados. La diferencia básica es la forma y esquema que siguen. Los datos estructurados tienen una forma y siguen un esquema fijo, lo que quiere decir que no hay que manipularlos para obtener información, y están enfocados a ser usados y almacenados en bases de datos convencionales. Los datos semiestructurados y no estructurados son aquellos obtenidos en forma denominada “bruta”. Suelen ser datos binarios, que no siguen ningún esquema ni forma y por tanto para poder analizarlos hay que tratarlos previamente (Camargo, J. 2014).

Un ejemplo de datos estructurados común sería un DNI (ocho dígitos y una letra), o una cuenta bancaria (20 dígitos) o una fecha de nacimiento (dd/mm/aaaa). Un ejemplo de dato no estructurado pueden ser un video, un SMS o una fotografía.

Vamos a analizar más en detalle esta clasificación:

- a) **Datos estructurados:** son todos aquellos que son generados con un esquema fijo y que tienen una estructura óptima para su almacenaje en bases de datos

convencionales (Powerdata, 2015). Dentro de los datos estructurados podemos encontrar cuatro subcategorías:

- Creados: quiere decir que han sido generados por los sistemas con un esquema predefinido.
- Provocados: al igual que los creados tiene un esquema predefinido, pero estos han aparecido por una acción indirecta llevada a cabo por una persona.
- Dirigido por transacciones: datos que se generan automáticamente cuando se cumple una acción de manera correcta.
- Compilados: son resúmenes de datos de interés tanto privados como públicos.
- Experimentales: son datos generados cuando se realizan estudios o experimentos.

b) **Datos semiestructurados:** estos datos son los que pueden estar en base de datos convencionales, aunque no tienen un esquema definido como para que su comprensión sea rápida y sencilla, sí tienen una serie de etiquetas que posibilitan el separado de un dato de otro (DataPrix, 2014). Hay que realizar una serie de acciones para poder adecuar su comprensión. Este tipo de datos híbridos son característicos del e-commerce, datos de estaciones meteorológicas o mercados emergentes.

Un ejemplo de datos semiestructurados es una tabla en la que se indiquen nombre, apellidos, correo y DNI de una persona. A esta tabla se puede añadir información que no tengan la misma estructura que los ya existentes.

c) **Datos no estructurados:** son todos aquellos datos que por sus características y esquema no pueden ser almacenados en una base de datos convencional y por ello tienen que ser tratados para poder extraer información de ellos. Hay dos subcategorías (Camargo, J 2014):

- Capturados: son los datos que se han creado a partir del comportamiento de un usuario.

- Generados por usuarios: son datos que el usuario origina específicamente como puede ser una publicación en una red social.

2.4.2. Origen.

Al principio, como es normal, los datos más importantes eran los datos estructurados, ya que no se poseía la tecnología suficiente como para comprender datos que no pudieran ser almacenados y procesados con bases de datos convencionales. Pero el auge de la tecnología ha jugado un papel fundamental en la creación de herramientas para poder analizar datos no estructurados, y esto no habría sucedido de no ser por la creación de todo tipo de objetos, sensores y tecnología en general que genera miles de millones de datos diariamente (Mayer, V 2013). Este análisis nos lleva a la clasificación de los datos según su origen.

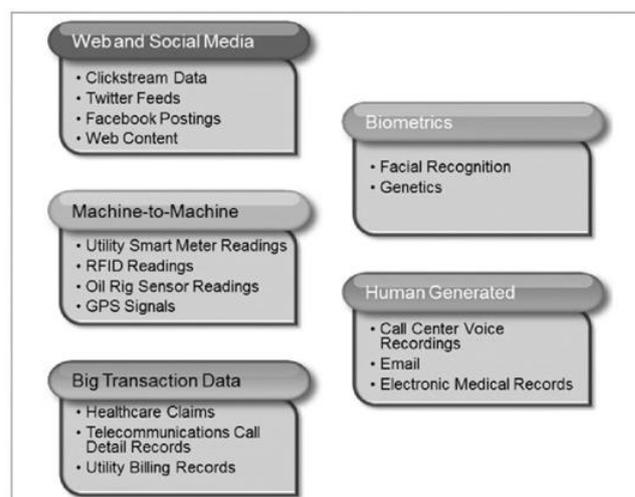


Figura 2: Ejemplos según origen de datos. Fuente: sofocorp.com (2014)

- a) **Datos web:** dentro de esta categoría de datos según su origen se incluyen todas las acciones que las personas llevamos a cabo dentro del mundo web, tales como número de clics, número de visitas en una web, búsquedas de información en Google, etc. Con una población con acceso a internet de 3.500 millones de personas, la cantidad de datos generada diariamente en esta categoría es enorme. En la ilustración 3 vemos la evolución del acceso a internet de la población mundial. Como podemos observar, el aumento de la población mundial con acceso a internet es uno de los causantes del aumento de la generación de datos cada año.

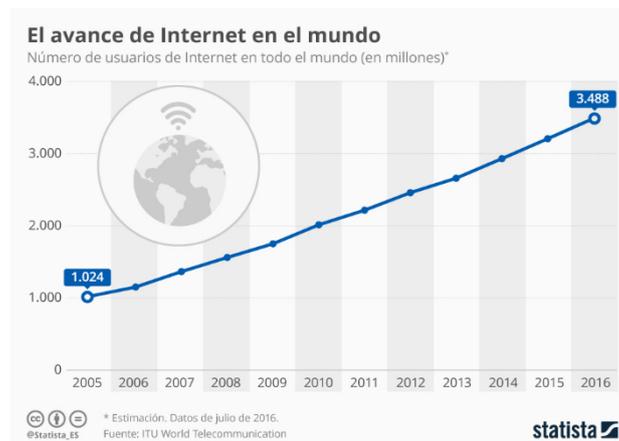


Figura 3: Fuente: Moreno, G (2016). Población con acceso a internet

b) **Datos de redes sociales:** probablemente son los datos más importantes basándonos en su cantidad hoy en día. Gran parte de la población con acceso a internet es usuario de una o más redes sociales, si anteriormente decíamos que 3,5 mil millones de personas tienen acceso a internet (We Are Social y Hootsuite, 2017) de estos, son 2,7 mil millones los que tienen perfil en una red social.

En la ilustración 4 podemos ver el número de horas semanales que empleamos en cada red social según Moreno, G (2017).

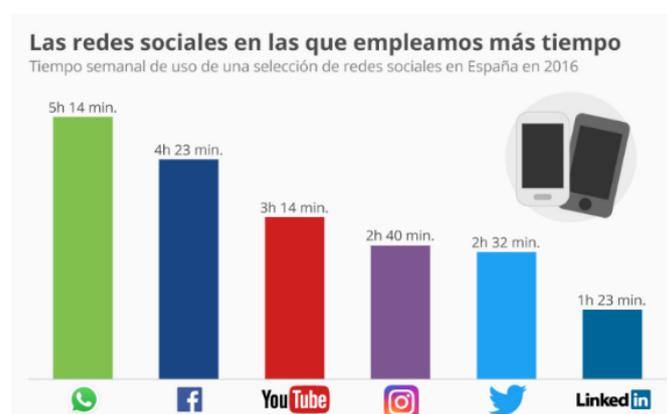


Figura 4: Fuente: Moreno, G (2017). Las redes sociales en la que empleamos más tiempo.

Vamos a destacar algunos datos sobre Facebook para así lograr hacernos una idea de la cantidad de datos que se pueden generar: El 22.9% de la población mundial

tiene Facebook; se suben 300 millones de fotografías diarias a esta red social; los usuarios pasan 20 minutos de media en Facebook cada día. Durante estos 20 minutos, el usuario ve perfiles, tanto personales como comerciales, da likes, deja comentarios o comparte publicaciones; en esos 20 minutos una persona genera datos que Facebook administra, estudia y vende a otras compañías. Gracias a ellos, Facebook está dando infinidad de soluciones para sus usuarios comerciales, información a marcas o perfiles comerciales por ejemplo cuándo publicar o frecuencia de aparición de sus publicaciones en su público objetivo.

Otro ejemplo es que con el envío de WhatsApp también se generan datos que la compañía vende a otras empresas para que lo puedan utilizar, pero hay que hacer una objeción, esto no significa que WhatsApp o la empresa a la que le venda los datos pueda leer nuestros mensajes, oír nuestros audios o ver nuestros videos o imágenes. Estos datos consisten no en la información que llevan dentro los mensajes sino en el número que mandas, horarios, a qué personas o los contactos que tienes números.

c) **Comunicaciones entre máquinas:** se refiere a los datos generados por la interacción entre objetos tecnológicos, cuyas siglas en ingles son M2M (“Machine to Machine) y se define como el intercambio de información y datos entre dos máquinas. Se estima que en 2020 habrá unos diez billones de máquinas M2M generando información. (Clemares, L, 2014). Los principales sensores que en la actualidad crean datos son:

- Sensores NFC: esta tecnología podemos definirla como un medio de emparejamiento entre máquinas para la transferencia de pequeños paquetes de datos (Pino, D. 2013)
- Señal GPS: este tipo de tecnología lleva ya varios años funcionando y se basa principalmente en la geolocalización de un objeto a través del intercambio de información de una máquina en la tierra con otra que se encuentra en el espacio (satélite)
- Otros sensores instalados en máquinas como parquímetros o máquinas expendedoras.

- d) **Transacciones:** se refiere al registro de diferentes acciones que realiza una persona como pueden ser llamadas, mensajería, registros de facturación con pagos online o pagos con tarjeta.
- e) **Biométricos:** actualmente se están generando nuevas tecnologías para el reconocimiento personal a través del reconocimiento facial, tecnología que por ejemplo ya integra el móvil iPhone X y por la que podemos desbloquear el móvil o pagar con tarjeta simplemente con el reconocimiento de nuestra cara. Todos estos reconocimientos se almacenan para posteriormente ser analizados. Otro ejemplo, la información del ADN también genera datos que pueden ser utilizados por hospitales o empresas farmacéuticas para desarrollar nuevos fármacos o nuevas técnicas hospitalarias.
- f) **Generados por personas:** en este caso nos referimos por ejemplo a las grabaciones de voz durante una conversación con teleoperadores de atención al cliente, correos electrónicos o los registros médicos de entidades públicas o privadas.
- Para hacernos una idea más general de los datos que se generan, en la ilustración 5 presentamos lo que sucede en internet en tan solo 60 segundos según Excelacom (2016).

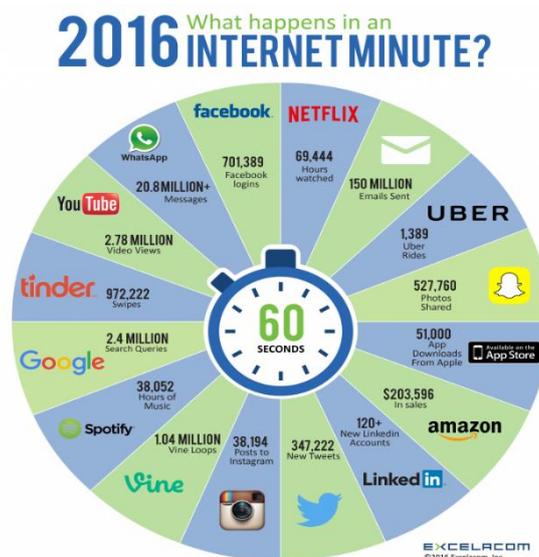


Figura 5: Fuente: Excelacom (2016). ¿qué sucede en internet en un minuto?

2.4.3. Almacenamiento.

La tecnología de almacenamiento ha cambiado en los últimos años, adaptándose al crecimiento de la necesidad de almacenar cada vez más datos. Todo este incremento y acumulación de datos ha afectado a su almacenamiento. El almacenamiento es una de las partes más importante de todo el proceso que comprende el Big Data y esto se debe a que el aumento de la cantidad de datos generada cada día se hace imprescindible por almacenarlos para así analizarlos posteriormente sin que se produzca una pérdida de información que puede ser muy valiosa para la institución.

Este proceso se basa en el “guardado” de los datos en forma bruta para que posteriormente puedan ser tratados y analizados. Las antiguas bases de datos relacionales han dado paso a potentes programas capaces de almacenar todo tipo de datos, sin importar tanto cuál es su estructura o su tamaño.

Podemos señalar que existen tres grandes tipos de almacenamiento utilizados en Big Data (PowerData.com, 2016). Estos son: NoSQL, “en memoria” y MPP. La utilización de uno u otro viene dada por las necesidades que pueda tener el usuario sobre los datos en cuestión.

a) **NoSQL:** es una clase de gestión de datos que difiere de los sistemas tradicionales de gestión de bases de datos, principalmente, en que no utiliza el lenguaje SQL para realizar las consultas. Como decíamos anteriormente, este tipo de almacenamiento no requiere de datos con una estructura fija y utiliza diferentes formas de almacenar datos como son clave-valor, formatos documentales o grafos. Vamos a describir cada uno de ellos:

- Clave-Valor: en los sistemas clave-valor tenemos contenedores en los que en su interior hay tantas parejas clave-valor como queramos. A cada clave, como puede ser NOMBRE se le asigna un valor dado. Un ejemplo sería NOMBRE=PEPE.
- Formato documental: está constituido por un conjunto de contenedores como la clave-valor, pero en su interior se encuentran documentos que no tienen estructura. Dentro de un mismo contenedor se puede almacenar tantos documentos como se quiera.



Figura 6: ejemplo de formato documental. Fuente:JorgeSanchez.net (2014)

- **Grafos:** en este caso los datos están almacenados en forma de grafos en los que se cuenta con nodos (entidades) y aristas (relaciones). Un nodo es un registro único que tiene al menos una propiedad. Las aristas por su parte definen las relaciones entre los nodos. Por ello los nodos pueden tener múltiples aristas que definan la relación que hay entre ellos. Este tipo de almacenamiento es útil ya que permite utilizar el análisis la teoría de grafos.

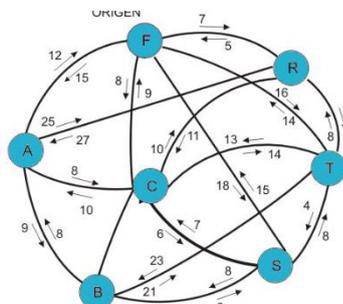


Figura 7: Ejemplo de grafo. Fuente: WordPress (2015)

El tipo de base de almacenamiento NoSQL, empezó a ser utilizada por las grandes empresas de internet, como son Facebook o Twitter entre otras, por la necesidad de almacenar, gestionar y analizar en vivo las grandes cantidades de datos que se generaban en sus webs cada minuto.

- b) **“En memoria”**: estas bases de datos tienen como principal característica que los datos están almacenados en su memoria principal para agilizar su utilización y obtener tiempos de respuesta rápidos. Esta base de datos está basada en la forma en la que lo hacen los discos duros o las memorias RAM, en los que la forma de acceder a los datos es rápida. Este tipo de sistema permite a los usuarios ejecutar consultas e informes sobre la información contenida. Si se trata de una organización conectada, la información generada y almacenada en los sistemas se actualizará automáticamente disminuyendo los tiempos de análisis por parte de los usuarios.
- c) **MPP** (Massiel Parallel Processor) es un sistema de proceso masivo en paralelo divide en muchas partes un conjunto de datos y lo distribuye como piezas independientes. La principal ventaja que tiene este tipo de base de datos es que puedes almacenar una gran cantidad de datos, trocearlos y tenerlos en servidores disponibles que no implique un coste muy alto.

3. Casos de éxito en la aplicación de Big Data.

A continuación, y con el fin de exponer la gran versatilidad y utilidad del Big Data, se van a exponer algunos casos de empresas que creemos que lo han utilizado con éxito.

1. Netflix.

Netflix es una empresa de Estados Unidos que se dedica a ofrecer un gran catálogo de series y películas para visionado online, cuando quieres, donde quieres y desde cualquier dispositivo a cambio de una cuota mensual que varía en función de si desea que haya más o menos dispositivos en los que se pueda ver un contenido a la vez (Boscan, S. 2017).

Esta empresa ha utilizado el Big Data en diversidad de áreas. A modo de ejemplo se han usado los datos recopilados en cada visionado para ofrecer un repertorio de películas y series que son del estilo y que le gusta ver a cada cliente, reduciendo el tiempo que el cliente tarda en buscar una serie o película que pueda ser de su gusto dentro de la gran cantidad de posibilidades que hay en Netflix (Boscan, S. 2017).

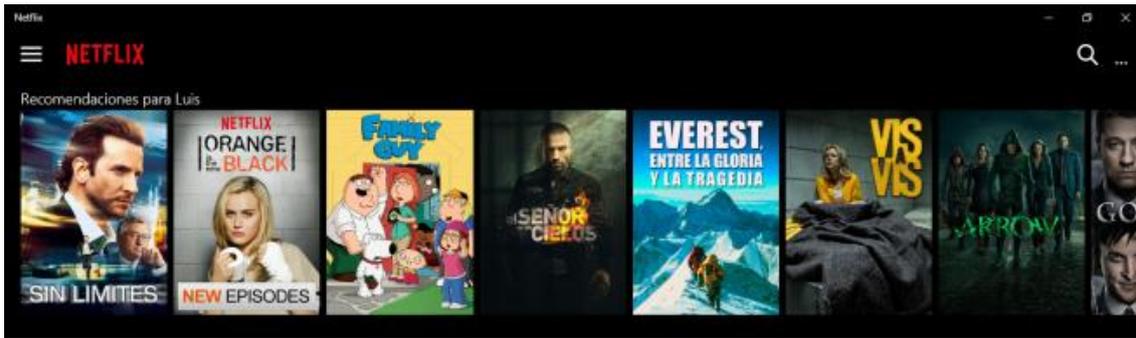


Figura 8: Recomendaciones de Netflix. Fuente: Captura de pantalla Netflix.com (2018).

Como vemos en la ilustración anterior, Netflix nos sugiere recomendaciones en función de las series y películas que ya hemos visto.

Otro caso en la utilización de Big Data por parte de esta empresa fue la creación de una serie muy conocida a partir de la recopilación y el estudio de los gustos y preferencias en series y películas de millones de usuarios. Lo que más demandaban, en este sentido, los usuarios eran contenidos que incluyeran drama, poder, política y sensualidad. Todas estas características se reunieron para crear la serie House Of Cards, de su propiedad y que ha tenido una repercusión enorme en sus ingresos. (Boscan, S. 2017)

2. Amazon.

Amazon es la mayor empresa de comercio online que hay actualmente en el mundo. Cuenta con millones de productos que pueden ser enviados a los compradores en menos de 24 horas y en algunas ciudades como Madrid o Barcelona este tiempo se reduce a un par de horas. Para comprar en Amazon no es necesario pagar una cuota, pero para beneficiarse de los envíos express o algunas utilidades más que ofrece si es necesario pagar una cuota anual (Amazon Premium).

Amazon, al igual que Netflix, ha utilizado el Big Data para mejorar la experiencia del comprador, logrando así más beneficios. En esta ocasión lo que se ha hecho ha sido ofrecerle una recomendación con cada producto que visita, basada en los miles de millones de datos recopilados a lo largo de todo el mundo. Otro claro ejemplo de la utilización del Big Data por parte de Amazon ha sido la presentación de

propuestas basadas en tus búsquedas de productos, así puede mostrar por ejemplo un libro que resulte ser de su interés antes de que lo busques. (Boscan, S. 2017)



Figura 9: Recomendaciones Amazon. Fuente: Captura de pantalla aplicación móvil Amazon (2018)

3. Nike.

Nike, una de las empresas de ropa y calzado deportivo más grande del mundo, también se ha unido a la utilización del Big Data. En este caso lo ha hecho a través de la recopilación de datos con sus dispositivos wearables³, como son los smartwachs o las zapatillas inteligentes, para ofertar a los usuarios productos concretos en función de la actividad que realizan (Post, R.2014).

4. T-Mobile.

T-Mobile es una de las grandes empresas de telecomunicaciones del mundo que gracias a la ayuda del Big Data ha logrado reducir el número de portabilidades⁴ que recibía anualmente en casi un 50%. Esto lo ha conseguido recopilando datos acerca de las quejas formales y a través de redes sociales que emitían sus clientes,

³ Wearable hace referencia al conjunto de aparatos y dispositivos electrónicos que se incorporan en alguna parte de nuestro cuerpo interactuando de forma continua con el usuario y con otros dispositivos con la finalidad de realizar alguna función concreta

⁴ Es el proceso que permite mantener un número de teléfono cuando se cambia de operador (Vodafone. 2018)

realizando acciones correctoras antes de que se hiciera más grande el problema. (Boscan, S. 2017)

5. Campañas electorales.

La utilización del Big Data no es exclusiva de empresas. El equipo de Barack Obama (Ex presidente de los Estados Unidos de América) decidió utilizarlo en su campaña de reelección de 2012. A través del uso del Big Data su campaña se centró en tres puntos claves: registro de los votantes que ya estaban decididos a votar a Obama, persuasión del electorado indeciso (suele ser un porcentaje muy importante en las elecciones americanas) y conseguir que los partidarios de Obama fueran a votar (Méndez, F. 2015).

El equipo de Obama consiguió que la inversión en publicidad de su candidato fuera más rentable gracias al cruce de datos de los votantes con otros demográficos, de audiencias o de precios de la publicidad. El impacto de la publicidad en el electorado mejoró sensiblemente con respecto a la campaña anterior (Méndez, F. 2015).

3.1. Utilización del Big Data en la educación internacional.

Dado que este trabajo se centra en la posible utilización del Big Data por parte de la Universidad Politécnica de Cartagena, se van a exponer en este epígrafe casos contrastados de éxito en la utilización de los datos por parte de universidades de otros países.

a) Escuelas de Innovación AltSchool de San Francisco.

Se trata de cuatro pequeñas escuelas ubicadas en los alrededores de San Francisco que se dedican principalmente al estudio de la innovación. En estas escuelas se está estudiando la posible utilización de los datos con el fin de mejorar la experiencia de aprendizaje por parte de los alumnos (More, M. 2015).

Para que el estudio sea lo más veraz posible se ha implantado una aplicación para todos los alumnos en la que deben ir indicando las actividades que realizan a lo largo del día a modo de ejemplo realizan un “check-in” a la hora que entran. Esta app se utiliza también como elemento de aprendizaje ya que alberga una playlist de actividades seleccionada por los profesores en función de los objetivos personales

de cada alumno. Los datos de progreso de cada alumno son almacenados para su posterior análisis por parte del profesor (More, M. 2015).

La idea es acumular todos estos datos con el fin de que los profesores puedan diseñar clases más efectivas y personalizadas.

b) Universidades que utilizan el Big Data como forma de evitar el abandono de los estudiantes.

En el ámbito universitario inglés el estudio realizado por Higher Education Funding Council for England (2016) destaca que un 8% de los alumnos que comienzan sus estudios universitarios lo abandona en el primer año.

Algunas universidades, atendiendo a este grave problema de pérdida de conocimiento humano que pueda ayudar al desarrollo socioeconómico del país, están empezando a utilizar los datos para prever la deserción en sus campus (Universia, 2016).

Estas universidades indican, utilizando para ello el análisis de Big Data, que la participación y presencia del estudiante es fundamental para evitar que deje la universidad. Una conclusión importante a la que han llegado es que sería conveniente no centrarse únicamente en factores étnicos o económicos a la hora de asignar las becas, sino también si estas son suficientes para cubrir todas las necesidades del estudiante durante su etapa universitaria (Universia, 2016).

La Ball Statu University de Indiana (EE. UU) ha dado un paso más allá y no se ha conformado con el análisis de los datos de compromiso académico, sino que también analiza sus actividades sociales con el fin de identificar cambios inesperados en los patrones de los estudiantes que puedan provocar dificultades en sus estudios.

Por otra parte, la Universidad de Arizona ha implementado un sistema en el que se avisa al estudiante cuando su rendimiento está bajando, dejándolo rezagado del grupo. Desde su implementación, en 2007, se ha aumentado en la tasa de titulación de un 26% a un 41% (Universia, 2016).

Estas universidades, al igual que el caso anterior de AltSchool, han llegado a la conclusión que una mejora en la forma de aprendizaje de los alumnos ofreciendo una experiencia más personalizada puede evitar en gran medida que estos abandonen.

c) Mejora en el programa de Aula Invertida.

El Aula Invertida es una metodología en la que los alumnos ven videos y contenido virtual en sus casas antes de desarrollar clases de carácter práctico presenciales (Sanz, I. 2017).

Con el análisis de los datos obtenidos del software de gestión de contenidos y reproducción de videos los profesores pueden mejorar la carga de trabajo en casa y en las clases presenciales, analizar los intereses de los alumnos de visionado de videos viendo los más vistos y los que no ven o evaluar la interacción entre los alumnos para preparar grupos de trabajo (Sanz, I. 2017).

Este ejemplo también podemos encontrarlo en la asignatura de Recursos Humanos, que se imparte en la Facultad de Ciencias de la Empresa de la Universidad Politécnica de Cartagena.

d) Universidad de Nottingham Trent.

Como otras universidades, la Universidad de Nottingham Trent le está dando mucha importancia a evitar el absentismo y la deserción de sus alumnos, para ello está llevando a cabo un programa piloto con el objetivo de anticiparse a una posible bajada de la motivación del alumno. Para poder dar una respuesta rápida a este problema la universidad comenzó a recolectar información de cuatro actividades que según ellos indicaban el nivel de compromiso del alumno, como son: el uso de la biblioteca, la cantidad y tipo de edificios de la universidad que visitan, la frecuencia de uso de las plataformas de educación y la cantidad de tareas y ejercicios entregados por medios electrónicos.

Según el departamento de la universidad encargado de analizar todos estos datos (Student Engagement Department, 2015), medir el nivel de compromiso de los estudiantes en estas cuatro áreas permite saber la probabilidad de éxito del mismo para pasar un curso o incluso de terminar la carrera y es que, según dicho departamento, solo el 25% de los estudiantes con un nivel de motivación bajo son capaces de superar el primer curso. Por ello, cuando se detecta un caso de motivación baja entre los alumnos automáticamente el caso se traslada a un tutor que se pone en contacto con ese alumno para entender porque tiene esos niveles tan bajos de motivación y poder darle una solución (Universia, 2016).

e) Universidad de Georgia.

Esta universidad es pionera en el uso de los datos ya que desde 2012 tiene implantado un software que cruza una gran variedad de estos que provienen de sus alumnos, previendo un posible estancamiento en la carrera. Este software está automatizado para que mande de forma autónoma correos electrónicos a los alumnos que detecta que tiene niveles de motivación bajos, indicándoles que deben acercarse a hablar con un tutor para que la situación pueda revertirse (Universia, 2016).

4. Análisis de la situación de las universidades españolas.

En este apartado, y con el fin de mostrar una idea de cómo se encuentra la situación de las universidades, tanto públicas como privadas, en España, vamos a hacer una breve exposición de qué es una universidad, las normas por las que se rige y la situación en la que se encuentran actualmente.

4.1. Universidad Pública en España.

Una universidad pública es aquella institución, en este caso de carácter público, aunque existen también privadas, que tiene como objetivo dar el servicio de enseñanza de estudios superiores como son los grados, máster y postgrados a la ciudadanía en general, sin distinción alguna, aunque sí con ciertas barreras de entrada como pueden ser las notas de corte que hay para cada carrera.

Estas instituciones se rigen por la Ley Orgánica de Universidades 6/2001 del 21 de diciembre que posteriormente fue modificada por la Ley 4/2007 del 12 de abril y en la que se especifican las normas y derechos por las que se rigen estudiantes, docentes y personal de las universidades tanto públicas como privadas, así como las principales funciones que deben desempeñar las universidades públicas en beneficio de la ciudadanía en general.

En la actualidad existen en España 50 universidades públicas distribuidas a lo largo de la geografía española, siendo las comunidades que más universidades tienen, Madrid, Cataluña, Valencia y Andalucía (MECD, 2017).

Comunidades y Ciudades Autónomas, Universidades Públicas, Universidades Presenciales, Con actividad, 2017-2018



Figura 10: Número de universidades públicas por CCAA. Fuente: MECD, (2018)

Existe una proporción de 1.73 universidades por cada millón de personas, y si tuviésemos en cuenta solo las personas que están en edad de estudios universitarios (entendiendo esta edad como la comprendida entre 18 y 24 años) esta cifra aumentaría hasta las 25.67 universidades por cada millón de persona (Ranking Universidad, 2017).

Con este último dato podemos hacernos una idea de la facilidad que existe actualmente para encontrar una universidad en la que estudiar que no esté muy lejos del lugar de residencia del alumno.

Motivado por esto y por la crisis económica de los últimos años, el número de alumnos matriculados en 2017 en alguna de estas 50 universidades ha sido de 1.545.732 (MECD, 2017). La población total que se encuentra en ese rango de edad hoy día en España es 2.261.020, lo que quiere decir que el 67.78% de la población en edad de realizar estudios universitarios los realiza (Universia, 2017).



Figura 11: Ministerio de Educación y Formación Profesional. Fuente: MECD, (2018)

Dentro de las universidades existen los denominados campus. Estos son centros donde se imparten los grados ofrecidos por la universidad. Una universidad puede tener varios campus; a modo de ejemplo, señalar que la UPCT tiene 4 campus, en donde están distribuidos los grados. Los campus se pueden ubicar en la misma ciudad que la universidad o en otra distinta, un ejemplo de ello es la Universidad de Valencia, que cuenta con un campus en Elche.

Cada vez es mayor el número de grados que se imparten en las universidades públicas españolas, este aumento viene por la necesidad de los alumnos de enfocar sus carreras a las nuevas necesidades del mundo laboral. El total de grados impartidos en España fue el año pasado de 2.854 (MECD, 2018). Sumado a estos estudios las universidades han empezado a implantar en su cartera de servicios otros complementarios como pueden ser los cursos para ser expertos universitarios, másteres oficiales y los másteres propios o en colaboración con centros privados.



Figura 11: Titulaciones impartidas 2017-2018. Fuente: MECD, (2018)

4.2. Universidad privada.

Podemos decir que un centro universitario privado es una organización (empresa) creada por personas físicas o jurídicas en virtud del apartado 6 del artículo 27 de la Constitución Española, que tiene como objetivo la obtención de beneficio a través de la realización de la actividad docente superior y la expedición de títulos homologados por los organismos competentes.

Las Universidades privadas se regirán por la Ley 4/2007, del 12 de abril, de Universidades, así como las normas dictadas por el Estado y las Comunidades Autónomas. También estarán sometidas a sus propias normas de organización y funcionamiento (Mela, M. 2012).

Según la Ley 4/2007 del 12 de abril, será la propia universidad privada la que decida cuál será su organización y funcionamiento, así como la forma en la que se distribuirán sus órganos de dirección y representación, la forma en la que se eligen, guardando siempre la correcta representación de todos los sectores de la comunidad universitaria.

Las principales ventajas que tiene la universidad privada según Martínez, A (2018) son las siguientes:

- La universidad privada contará con mayores recursos tanto económicos como tecnológicos con respecto a la pública ya que no le afectarán los posibles recortes presupuestarios por parte del gobierno ni los conflictos políticos.
- Tendrá unos planes de estudios más específicos ya que al no depender del estado tiene mayor flexibilidad a la hora de elaborar los mismos y pueden adaptarlo mejor a las necesidades de los futuros alumnos.
- En muchas ocasiones realizar los estudios universitarios en una universidad privada tiene mayor prestigio que hacerlo en una pública, esto está ocasionado por los diferentes rankings anuales en los que de las 10 primeras universidades que mayor prestigio tienen en nuestro país 7 son privadas (Ranking Universia, 2017)

4.3. Universidad pública vs universidad privada.

Este trabajo está enfocado a las posibles aplicaciones que tiene el Big Data en la universidad pública y su posible utilización como forma de captar nuevos alumnos, por este motivo consideramos importante establecer el contexto actual existente entre la universidad pública y la privada.

Es en la obtención de beneficio donde radica la mayor diferencia entre una universidad y otra. Si bien la universidad pública no tiene como objetivo la obtención de beneficio, ya que es verdad que el coste que asume el estado por cada alumno es mucho mayor que el que se le repercute a través de la matrícula. Según la Fundación Conocimiento y Desarrollo (2015) y su informe anual, el coste medio de un alumno para el estado es de 6.480€ y por el contrario el coste medio de una matrícula universitaria es de 1.564€, sí que necesita de un número mínimo de alumnos por cada estudio para que sea “rentable”. Con la universidad privada

sucede algo parecido, es necesario un número de alumnos mínimo para que la actividad sea viable para los empresarios que están llevando a cabo esa actividad. Tenemos el producto, que son los diferentes grados que se imparten en las universidades y los números de las universidades públicas los hemos estudiado en el punto anterior. De las universidades privadas sabemos que hay un total de 33 repartidas por España y que en ellas se imparten un total de 562 grados académicos. Esto significa que casi 4 de cada 10 universidades que hay en España es privada, número que se espera que aumente en los próximos años (Universia, 2017). En la siguiente imagen se muestra la distribución de las universidades privadas por Comunidad Autónoma

Comunidades y Ciudades Autónomas, Universidades Privadas, Universidades Presenciales, Con actividad, 2017-2018



Figura 13: Universidades privadas por CCAA. Fuente: MECD, (2018)

Una vez estudiado el producto tenemos que estudiar más el cliente potencial, los alumnos, y sobre todo la forma de llegar a ellos. Es aquí donde compiten los dos tipos de universidades, tanto unas como otras quieren cada año más alumnos en detrimento de las demás. La forma que tienen las universidades para obtener cada vez más alumnos, además de aumentar su prestigio, es la realización de campañas de captación a través de todos los medios que estén a su alcance.

Las universidades privadas cuentan con una mayor inversión destinada a captación de nuevos alumnos que la que puede realizar una universidad pública ya que ésta tiene un control burocrático sobre en qué se utiliza el dinero que ralentiza y paraliza, en algunas ocasiones, la iniciativa de los gestores universitarios.

En definitiva, la universidad pública cuenta con un prestigio mayor que la privada por el número de años que lleva activa, pero la universidad privada puede hacer

más inversión en captación y ya que el prestigio se adquiere con los años, pensamos que llegará un momento que, si la universidad pública no cambia su política de captación, la privada ganará más alumnos y por tanto crecerá más.

5. Posible implantación en la Universidad Politécnica de Cartagena.

Existen ciertos problemas en la universidad pública española en general sobre la forma de actuar a la hora de tratar el absentismo y el desánimo de los estudiantes, así como la forma de enseñanza educativa actual. Estos problemas se traducen en un nivel de abandono cada vez mayor, pasando del 19% del periodo 2011/2012 al 22,5% del periodo 2015/2016 (Fundación Conocimiento y Desarrollo, 2017), o en el propio desánimo de los profesores por la dedicación en sus asignaturas.

Como mencionábamos, la finalidad de este Trabajo Fin de Estudios es realizar alguna aportación para la posible implantación en la Universidad Politécnica de Cartagena del uso del Big Data, obteniendo un feedback de los propios alumnos y dando una posible solución así a los problemas planteados.

Para poder exponer esta idea vamos a analizar las herramientas con las que cuenta actualmente la universidad, la legislación por la que se rige el uso de datos tanto en el seno estatal como en el comunitario y qué podría suponer la implantación del uso de dichos datos para la universidad.

5.1. Situación y herramientas actuales.

Actualmente la Universidad Politécnica de Cartagena cuenta con un departamento dedicado al análisis de datos, la Oficina de Prospección y Análisis de Datos (OPADA). Este departamento fue creado en junio de 2014 (-301/14 de 2 de junio) con el objetivo de ser la unidad técnica especializada en gestión de información de la universidad. Tiene como objetivos específicos (OPADA, 2014):

- Normalizar, sistematizar y facilitar la gestión de los datos mediante la utilización de herramientas informáticas avanzadas.
- Abordar la gestión integral de los datos relacionados con la UPCT, tanto de fuentes internas como externas para posteriormente transformarlas en valor añadido que ayuden a la dirección estratégica de la universidad.

- Fomentar la transparencia a la hora del tratamiento de datos.
- Buscar alianzas y colaborar con otros grupos de interés especializados en el análisis de datos.
- Hacer de la prospección un pilar fundamental del tratamiento de la información, como forma de ofrecer respuesta a la universidad y la ciudadanía en general.

Para poder llevar a cabo todos estos objetivos la OPADA cuenta con una serie de herramientas informáticas, entre las que podemos encontrar bases de datos y programas de inteligencia institucional.

Cuentan con un programa de base de datos objeto-relacional, “Oracle Database”. Se trata de una base de datos en cloud, es decir, que almacena los datos en servidores que no se encuentran físicamente en las instalaciones de la universidad. Este servicio esta prestado por la empresa de soluciones tecnológicas Oracle.

Para poder desarrollar las estrategias empresariales, que posteriormente se le harán llegar a la dirección de la universidad, cuentan con un conjunto de programas libres que generan inteligencia empresarial llamados Pentaho BI Suite.

Ya que manejan gran cantidad de datos, y que su visualización se haría imposible de forma convencional, han recurrido al software libre Data-Driven Documents para la manipulación de los datos. Para la mejor visualización de la información se apoyan en herramientas como Pentaho Reporting, que es un software de generación de informes y motor de presentación.

Durante los 4 años se han realizado diferentes actividades que OPADA ha ido detallando en informes de actividades anuales. A continuación, exponemos una pequeña muestra de estas actividades:

La secretaria de la ETSII solicita el listado de los alumnos de nuevo ingreso con nota de admisión, curso 2016-17.	jul-17
La Subdirección de la ETSIA solicita detalle de tesis y TFE en 2015-16.	jul-17
La subdirección de la ETSINO solicita datos relativos a las prácticas de alumnos en empresa de GANISM, curso 2015-16, para informe de la ANECA.	jul-17
El Vicerrectorado de Tecnologías de la Información y las Comunicaciones solicita el listado de matriculados en 2017-18 con nota de admisión mayor o igual a 11.	jul-17
El Vicerrectorado de Ordenación Académica y Calidad solicita detalle de las últimas asignaturas de grado que aprueban los estudiantes.	jul-17
El Servicio de Gestión de la Calidad solicita la generación de los informes de satisfacción de los estudiantes con la actividad docente, segundo cuatrimestre 2016-17.	jul-17
La Subdirección de la ETSIA solicita el listado de alumnos de la ETSIA participantes en programas de movilidad en 2015-16.	jul-17

Figura 14: Listado de tareas realizadas por la OPADA. Fuente: OPADA, 2018

5.2. Legislación Vigente.

Ya que los datos que maneja OPADA son de carácter personal, y la obtención y utilización de los mismos se ha de hacer según las leyes estatales y comunitarias vigentes, pasamos a hacer una relación de dichas normas.

5.2.1. Ley de Protección de Datos.

La Ley Orgánica 15/1999 de 13 de diciembre de Protección de Datos de Carácter Personal lleva activa desde el año 2000 y se fundamenta en el artículo 18 de la Constitución Española (Constitución Española):

“Se garantiza el derecho al honor, a la intimidad personal y familiar y a la propia imagen”

Recientemente, y motivado por el auge comentado de la recolección y análisis de datos, se ha modificado dicha Ley de Protección de Datos adaptándola a las nuevas circunstancias. Estos cambios se pueden resumir en los siguientes (Tomás, A. 2017):

- Los datos recogidos han de ser solo los necesarios para la actividad que se precisen.
- Hay que especificar qué datos van a recogerse y con qué fin se van a utilizar.



- Facilitar un modo de contacto para que el usuario pueda editar o cancelar los datos.
- Es necesario el consentimiento expreso de la cesión de los datos y estos pueden ser revocados.
- La organización que reciba, almacene y trate los datos ha de tener las medidas de ciberseguridad necesarias para el nivel de riesgo que tenga esos datos.
- El responsable de los datos tendrá que hacer llegar informes sobre los mismos a la Agencia de Protección de Datos.

Toda organización ya sea pública o privada ha de cumplir con todos estos requisitos para poder manejar datos personales. El no cumplir con las normativas expuestas puede conllevar diferentes sanciones dependiendo de la gravedad de la infracción cometida. El organismo que vela por el cumplimiento de la LOPD es la Agencia de Protección de Datos (APD), también encargada de definir y establecer los grados de gravedad, así como las sanciones. Actualmente la APD ha establecido tres grados de infracción, leve, grave y muy grave.

Se tratará de una infracción leve, por ejemplo, el no rellenar el formulario y depositarlo en la APD. Esta infracción conlleva sanciones económicas de entre 900 y 40.000€.

Una infracción grave, es por ejemplo obtener datos personales sin consentimiento, y tiene multas de 40.000€ a 300.000€.

Por su parte, una infracción muy grave es la obtención de los datos de manera ilegal y con engaños al usuario y conlleva multa de 300.000€ a 600.000€.

5.2.2. Reglamento General de Protección de Datos.

El Reglamento Europeo de Protección de Datos fue aprobado por el Parlamento Europeo en 2016, y ha sido en mayo de este año 2018 cuando ha entrado en vigor. “Con este reglamento de protección de datos conseguimos un nivel uniforme de protección en toda la UE. Es un gran éxito para el Parlamento y un “sí” claro a los derechos de los consumidores y a la competencia en la era digital. Los ciudadanos podrán decidir por sí mismos qué información quieren compartir”, subrayó tras la aprobación en una entrevista Jan Albrecht (Verdes, Alemania), responsable de la tramitación parlamentaria del texto (2016).

En el nuevo Reglamento aprobado se especifican las normas que deben cumplir todas las empresas a la hora de tratar los datos de los ciudadanos, así como los derechos que tienen las personas frente a las empresas. A continuación, se especifican los cambios más relevantes:

- Se amplía el territorio de aplicación, ya no sólo afectara a empresas europeas si no que será aplicado a empresas que vayan a tratar datos de ciudadanos europeos. Este cambio va destinado a la mejora de la protección de los ciudadanos europeos cuando existan transacciones en internet con empresas no europeas.
- Se obliga a las empresas a emitir nuevos mensajes sobre por ejemplo el tiempo de retención de los datos. Además, será obligatorio informar de manera clara y entendible por el usuario.
- Se prohíbe la obtención de datos más allá de los necesarios. Será necesario el consentimiento expreso y libre del usuario, con esto se evita que las empresas obtengan los datos por el silencio del usuario.
- La empresa que vaya a tratar los datos deberá de tener un registro de las actividades en el tratamiento de los datos.

Además de estos cambios, en el nuevo Reglamento Europeo se incluye el Derecho al Olvido, que ya dictara el Tribunal de Justicia Europeo en 2014 y por el que la empresa está obligada a eliminar los datos de personas que consideren que sus datos son obsoletos o no son relevantes.

5.3. ¿Qué aportaría y qué supondría el uso del Big Data para la UPCT?

La UPCT cuenta actualmente con un total de 1.243 alumnos repartidos en quince titulaciones de grado, veintidós titulaciones de máster y ocho doctorados (Portal Transparencia UPCT, 2018). Todos estos alumnos generan una cantidad enorme de datos que, como hemos visto, son parcialmente usados por un departamento de la universidad para poder mejorar o dar información a otros departamentos, ya sea para crear previsiones o definir nuevas estrategias.

Gracias a que la universidad ya cuenta con el departamento encargado de la recolección y análisis de los datos, el coste de ampliar las actividades a realizar por

este creemos que sería muy bajo. Consideramos que para hacer viable este departamento se podría llegar a acuerdos con empresas tanto regionales como de otras partes de España para conseguir mayor financiación para su desarrollo. Existen otros casos similares con proyectos realizados por la universidad en colaboración con empresas como puede ser el Banco de Santander.

Para conseguir que este departamento pueda ser un pilar muy importante para la universidad podría utilizarse como método para realizar becas o prácticas para los alumnos que estén interesados, aumentando el personal que trabaja en el mismo a un coste menor en comparación con contratar personal especializado. En la actualidad hay tres personas trabajando en él pero pensamos que aplicando las actividades que se van a ver a continuación el departamento podría llegar a dar trabajo a unas 10-15 personas, entre personal técnico encargado de comprobar que la recolección de los datos se hace de manera adecuada y de su posterior procesamiento, analítico, realizando los informes necesarios y dándole sentido a los datos, y de mantenimiento para tener siempre en perfecto estado todos los software y hardware necesarios.

A continuación, se detallan una serie de actividades que podrían ser de gran interés para la universidad y que pensamos supondría avances significativos:

1. Monitorización de la actividad de los alumnos en los diferentes portales web universitarios (Aula Virtual, Portal de Servicios, etc.)
2. Seguimiento de la actividad de los alumnos en las diferentes instalaciones físicas pertenecientes a la universidad (Bibliotecas, aulas de informática o centros deportivos)
3. Seguimiento de las evaluaciones a alto nivel por asignatura y notas medias de grado.
4. Seguimiento de la interacción de los alumnos con el resto de comunidad universitaria.
5. Control de absentismo a través de la participación a clase.
6. Monitorización de los datos de los alumnos que ya han finalizado sus estudios.
7. Recopilación de datos de alumnos pre-universitarios.

Las cinco primeras actividades se engloban dentro del llamado Learning Analytics que como dice Andres Nuñez (2017) “es el uso inteligente de los datos, derivados del comportamiento de los alumnos [...] como el número de accesos, los materiales revisados, las participaciones y evaluaciones y demás información que comprende valor para la institución universitaria”. Pensamos que podrían tener una serie de repercusiones positivas a medio/largo plazo tanto para el alumno como para la universidad.

Gracias al seguimiento de la actividad de los alumnos en los portales web, la universidad podría ver cuáles son más útiles, los de mayor uso y así como los que no están correctamente diseñados o no tienen una utilidad para el alumnado, pudiendo hacer encuestas para incorporar nuevas funcionalidades. Con el seguimiento de la actividad de cada alumno en la biblioteca, tanto física como virtual, creemos que se podría recomendar material bibliográfico a otros alumnos que tengan un perfil similar (De Mesa, M. y Gironda, J. 2018).

Un seguimiento exhaustivo de las calificaciones de las diferentes asignaturas y notas medias de los grados creemos que podría proporcionar al profesorado puntos de mejora dentro de sus asignaturas. Aparte de esto, pensamos que, con un conocimiento mayor de las preferencias de los alumnos en cuanto a las asignaturas impartidas, se pueden contemplar nuevas materias que estén más acordes con los estilos de enseñanzas actuales y lo que las empresas posteriormente van a demandar a los alumnos.

Para el control del absentismo se podrá utilizar el nuevo programa piloto implantado en la Facultad de Ciencias de la Empresa y en la Escuela de Telecomunicaciones denominado “Programa Tutor” por el cual, como comenta la profesora Eva Martínez Caro, (Noticias UPCT, 2018) “Cada uno de los alumnos de nuevo ingreso que se matriculen en la Escuela de Telecomunicaciones o la Facultad de Ciencias de la Empresa, contará con un tutor que se encargará de realizarle un seguimiento continuo y permanente”. Actualmente el programa piloto está integrado por quince alumnos de nuevo ingreso y quince tutores. Este programa también está implantado en la Universidad de Valencia bajo el nombre “Entreiguals, Mentoria d’estudiants”. A diferencia del programa Tutor de la UPCT, en la Universidad de Valencia son los

propios alumnos de cursos superiores los que realizan la acogida a los alumnos de nuevo ingreso o a los alumnos incoming⁵ (Univ. Valencia, 2018).

El “Programa Tutor” pensamos que podría servir como método de contacto entre la universidad y el alumno para informar a este último de tendencias que puedan provocar su ausencia a clase, bajada de notas o incluso el abandono de la carrera. Creemos que, al estar en contacto desde el primer día, es el tutor la persona más indicada para poder comunicar este tipo de información, de una forma lo más eficiente posible y que no cree en el alumno el efecto contrario,.

Con el seguimiento de los alumnos que ya han terminado sus estudios en la universidad, creemos que se podría conocer la tendencia que existe en cada año en cuanto a empleabilidad y que es lo que las empresas requieren en cada momento. Conociendo estos datos creemos que la universidad se vuelve más versátil y puede ofrecer a los alumnos la formación que más se requiere en el mundo laboral y proporcionarle mejores oportunidades de trabajo.

Creemos, que como ya consiguió la Universidad de Saint Louis (noticias. universia, 2018), la recolección y estudio de los datos de los alumnos de secundaria y bachillerato podría proporcionar a la universidad una información muy valiosa de cara a preparar nuevas estrategias de captación de nuevos alumnos, así como de la oferta académica acorde con lo que demandan.

Todos estos cambios supondrían una serie de beneficios:

- Disminución de costes por parte de la universidad, ya que pasaría a utilizar de manera más eficiente los recursos con los que cuenta.
- Disminución del absentismo del alumnado.
- Convertir la universidad convencional en una universidad disruptiva, en consonancia con los tiempos actuales. Creemos que puede convertir a la UPCT en una universidad mas innovadora, puntera y de calidad.
- Aumento del prestigio de la universidad por el aumento de la dedicación del alumnado.
- Captación de un mayor número de alumnos para próximos cursos.

⁵ Alumno incoming es aquel que viene de otra universidad, ya sea nacional o extranjera (Univ. Valencia, 2018)

- Pensamos que todos estos beneficios también se volcarían en los alumnos, creando graduados cada vez más preparados y más capacitados de enfrentarse a problemas en el mundo laboral actual.

6. Conclusiones.

La finalidad de este Trabajo Fin de Estudios ha sido plasmas la importancia que creemos debería tener una mayor utilización del Big Data en la Universidad Politécnica de Cartagena. Para poder llevar a cabo el objetivo primero dábamos a conocer en que consiste el Big Data y la evolución durante la última década, en la que se ha producido una gran explosión tanto en el análisis de los datos como en la cantidad y la velocidad de generación de estos.

Todos estos datos tienen una estructura y un origen diferente, pudiendo encontrarnos con datos estructurados, no estructurados, procedentes de redes sociales o biométricos. Es por esta variedad que su análisis se hace cada vez más imprescindible y por ello cada día se desarrollan nuevos softwares, cada vez más potentes y que son capaces de analizar cada vez más datos. En 2018 se crean 17 EB mensuales, y es por eso por lo que en la actualidad cobra gran importancia para la empresa el crear un método diferenciador del resto de competidores a través del Big Data.

Una vez conocido el Big Data y su historia hemos expuesto sus utilidades mostrando diferentes casos de éxito en la empresa privada como pueden ser Netflix o Amazon. También hemos podido conocer como el uso por parte de algunas universidades extranjeras ayuda a conocer mejor las experiencias de los alumnos, o prevenir el absentismo.

En el punto siguiente dábamos a conocer las características de las universidades públicas y privadas, así como las normas legales por las que se rigen cada una, lo que ha servido de nexo para dar paso al siguiente punto y poder conocer lo que se está realizando actualmente en la Universidad Politécnica de Cartagena en referencia al análisis de datos.

Dentro de este punto hemos conocido la Oficina de Prospección y Análisis de Datos (OPADA) que se encarga de realizar informes para los diferentes organismos universitarios con los datos de los que disponen. Posteriormente, creíamos necesario exponer la legislación vigente sobre protección de datos, y una vez conocido esto,

hemos expuesto las consecuencias que podría tener la utilización del Big Data por parte de la Universidad Politécnica de Cartagena.

En definitiva, creemos que el Big Data es una herramienta indispensable para cualquier organización que quiera crear una diferenciación sobre sus competidores y con el transcurso del tiempo y el avance de la tecnología se va a hacer cada vez más imprescindible para poder ofrecer al cliente una experiencia cada vez más personalizada.

El objetivo de este proyecto consistía en el estudio del Big Data como herramienta en auge y el análisis de una posible utilización por parte de las universidades tanto públicas como privadas, aunque centrando nuestros objetivos principalmente en la Universidad Politécnica de Cartagena. Así pues, tras dar una visión general del avance de este concepto, conocer las herramientas y departamentos existentes en la UPCT y tras un análisis que hemos realizado en los puntos detallados a lo largo del trabajo, creemos poder afirmar que el uso del Big Data por parte de las universidades aportaría al alumno una experiencia mucho más completa e individualizada. Esto es debido a que gracias a esta herramienta podrían ofrecer a los alumnos actividades más centradas en sus intereses, asignaturas más orientadas a los objetivos y exigencias que presentan los mercados de trabajo en la actualidad. Lo que permitiría a las entidades conocer las inquietudes del alumnado e incluso volverse más competitivas. ¿Por qué creemos que esto sería posible? Actualmente y desde no hace tanto tiempo, todos los organismos, tanto públicos como privados, tienen páginas web y perfiles en redes sociales debido al generalizado uso de Internet en la sociedad.

Este uso generalizado de Internet es lo que ha servido de impulso al Big Data, y lo que ha hecho que esta herramienta tenga hoy en día la importancia que está tomando.

Teniendo en cuenta las limitaciones que puede tener la Universidad Politécnica de Cartagena como organismo público para poder destinar fondos a este tipo de actividades, creemos que el uso del Big Data podría hacer a esta entidad mas competitiva y de mayor interés para alumnos tanto del territorio nacional como internacional. ¿Cómo podría hacerse esto posible?

Como hemos visto en nuestro trabajo, podemos acceder a gran cantidad de datos que proporcionamos diariamente con las consultas que hacemos en la web, el tiempo que permanecemos mirando una foto en Facebook o el número de mensajes

que enviamos diariamente en WhatsApp. Sólo tenemos que tener claro cuáles son nuestros objetivos y cuál sería nuestro público.

Cuando pensamos en el periodo universitario nos centramos básicamente en los jóvenes de entre 18 y 30 años que buscan completar una formación de calidad para su posterior ingreso en el mercado laboral. Todos los alumnos pasamos por los centros educativos pensando en el día de mañana ser grandes profesionales y tener un futuro prometedor dedicándonos a aquello que hemos elegido estudiar en nuestra universidad. Para que esto sea así, es necesario que las universidades planteen esta etapa de forma competitiva, es decir, nos muestren asignaturas atractivas, en las que se traten temas de actualidad y con las que podamos defendernos fuera cuando nos enfrentemos a una entrevista de trabajo o a mantener un puesto frente a otros compañeros, o nos proporcionen charlas y cursos orientados a temas de interés entre los alumnos de la universidad.

En nuestra opinión todo esto sería posible a través del uso del Big Data ya que podríamos conocer en profundidad los intereses del sector de población que es de nuestro interés y tomaríamos decisiones en función de los datos obtenidos, lo que además nos proporcionaría la certeza de que no son decisiones tomadas a la ligera, sino que están tomadas a raíz de un análisis de datos contrastados y por lo tanto tendrían un porcentaje elevado de éxito.

7. Bibliografía.

Adriana, M (2018). 10 ventajas de la universidad privada. Consultado en:
<https://www.avanzaentucarrera.com/llegaraser/orientacion-academica/como-elegir-grados-universitarios/10-ventajas-universidad-privada/>

Albrecht,F(2016). Reforma de la protección de datos – Nuevas reglas adaptadas a la era digital. Consultado en:
<http://www.europarl.europa.eu/news/es/headlines/priorities/20130901TST18405/20160407IPR21776/reforma-de-la-proteccion-de-datos-nuevas-reglas-adaptadas-a-la-era-digital>

Baranco, R (2012). ¿qué es big data? Consultado de:
<https://www.ibm.com/developerworks/ssa/local/im/que-es-big-data/>

Big data (2017). Consultado en web:
https://es.wikipedia.org/wiki/Big_data#cite_note-13

Big data una herramienta que utilizan las universidades para evitar la deserción (2016). en:

<http://noticias.universia.com.ar/educacion/noticia/2016/12/26/1147817/big-data-herramienta-utilizan-universidades-evitar-desercion.html>

Boscan, S (2017) Casos de éxito: el uso del big data. Consultado en: <https://unimooc.com/casos-exito-uso-del-big-data/>

Camargo, J (2014) Conociendo el big data. Colombia, Universidad Pedagogía y tecnología de Colombia

Cambios en la LOPD para 201 (2017). Consultado en: <https://factorialhr.es/blog/cambios-proteccion-datos-2018-lopd/>

clemares, L (2014) ¿Qué es la comunicación M2M? consultado en: https://www.tendencias21.net/telefonica/Que-es-la-comunicacion-M2M_a801.html

Constitución Española (2018). Consultado en: <http://www.congreso.es/consti/constitucion/indice/titulos/articulos.jsp?ini=18&tipo=2>

Ford and Nike use big data to make smarter sustainable design: <https://www.theguardian.com/sustainable-business/ford-nike-big-data-smart-sustainable-design>

Franceschin, T (2017) Big data en universidades. Consultado en: <http://edu4.me/big-data-universidades/>

Galimany, A (2014). La creación de valor en las empresas a través del Big Data. (Trabajo Fin de Grado). Universidad de Barcelona.

Informa CYD (2014). La contribución de las universidades españolas al desarrollo. Consultado en: <http://www.fundacioncyd.org/informe-cyd/informe-cyd-2014>

Joyanes, L. (2013). Análisis de grandes volúmenes de datos en organizaciones (1ra ed.). México, Alfaomega.

Lara, F (2017). Conoce la historia del big data. Consultado en: <https://cl.godaddy.com/blog/conoce-la-historia-del-big-data/>



Las 7 V del Big data: Características más importantes (2016) Consultado en:

<http://www.iic.uam.es/innovacion/big-data-caracteristicas-mas-importantes-7-v/#veracidad-datos>.

Las empresas, ante el reto de conseguir un impacto económico real con Big Data

(2017) Consultado en: <http://www.ituser.es/big-data/2017/10/las-empresas-ante-el-reto-de-conseguir-un-impacto-economico-real-con-big-data>

LEY ORGÁNICA 4/2007, de 12 de abril, de Universidades (2007). consultado en:

<https://www.boe.es/buscar/doc.php?id=BOE-A-2007-7786>

LEY ORGÁNICA 6/2001, de 21 de diciembre, de Universidades (2001).

consultado en: <https://www.boe.es/boe/dias/2001/12/24/pdfs/A49400-49425.pdf>

Mayer, V y Cukier, K (2013). Big Data. La revolución de los datos masivos. Madrid, Turner Publicaciones.

Méndez, F. (2015) ¿Cómo el Big Data ayudó a Obama a ganar?:

<http://forbes.es/emprendedores/7560/como-el-big-data-ayudo-a-obama-a-ganar/>

Ministerio de Educación y Formación Profesional, (2017). Estadísticas sobre el

curso 2017-2018. Consultado en: <http://www.mecd.gob.es/servicios-alcidadanomecd/estadisticas/educacion/universitaria.html>

Monero, G (2017). Facebook y WhatsApp, las redes con más uso en España.

Consultado en: <https://es.statista.com/grafico/7887/facebook-y-whatsapp-las-redes-con-mas-uso-en-espana/>

More, M (2015). Los beneficios del uso del big data. Consultado en:

<https://www.iebschool.com/blog/usar-big-data/>

Moreno, G. (2016). Un 44% de la población mundial tiene acceso a Internet.

Consultado el 30 de junio en: <https://es.statista.com/grafico/6791/un-44-de-la-poblacion-mundial-tiene-acceso-a-internet/>

Noticias UPCT (2018). Los alumnos de nuevo ingreso de la UPCT tendrán un tutor durante la carrera a partir de septiembre. Consultado en:

http://www.upct.es/destacados/cdestacados.php?c=6&ubicacion=general&id_buscar=9437



Núñez, A (2017) Esta es la herramienta 2.0 que está ingresando a las aulas:
<https://blog.andresnunez.com/esta-es-la-herramienta-2-0-que-esta-ingresando-a-las-aulas/>

Oficina de Prospección y Análisis de Datos (2018). Consultado en:
<http://www.upct.es/opada/es/presentacion/presentacion>

Oficina de Prospección y Análisis de Datos (2018). Consultado en:
http://www.upct.es/gestionserv/inter/web_servopada_fich/doc_secciones/240actividades-2016-17.pdf

Otto, C (2016). ¿es legal que Facebook y WhatsApp compartan tus datos?

Consultado en:

<http://www.lavanguardia.com/tecnologia/20161009/41843913284/facebook-whatsapp-compartan-datos.html>

Pino, D (2013). NFC, qué es y cómo funciona. Consultado en:
<https://www.tuexpertomovil.com/2013/05/24/nfc-que-es-y-como-funciona/>

Sanz Isabel (2017) Infografía: cómo utilizar el big data en el aula invertida.

Consultado en: <https://www.u-planner.com/es/blog/infografia-como-aprovechar-al-maximo-el-big-data-en-el-aula-invertida>

Vidal, J (2012). Big Data: Gestión de datos no estructurados. Consultado en:

<http://www.dataprix.com/blog-it/big-data/big-data-gestion-datos-no-estructurados>