



UNIVERSIDAD DE JAÉN
Facultad de Ciencias Sociales y Jurídicas

ESTUDIO DE LAS POSIBILIDADES DEL USO DE TÉCNICAS DE ANÁLISIS DE BIG DATA PARA PYMES

Alumno: María Puerta Sánchez

Julio, 2017



UNIVERSIDAD DE JAÉN
Facultad de Ciencias Sociales y Jurídicas

Don Carlos Molina Fernández, tutor del Trabajo de Fin de Grado titulado: Estudio de las posibilidades del uso de técnicas de análisis de Big Data para Pymes, que presenta María Puerta Sánchez, autoriza su presentación para defensa y evaluación en la Facultad de Ciencias Sociales y Jurídicas de la Universidad de Jaén.

Jaén a 26 de Junio de 2017.

El alumno:

El profesor:

Fdo.: María Puerta Sánchez

Fdo.: Carlos Molina Fernández

*A **mi pareja**, Javier, por las horas invertidas, por todos los viajes perdidos, por los días libres no disfrutados, por los fines de semana trabajados, por el gran conocimiento que me ha transmitido, por escuchar y apoyar mis ideas, por ser mi compañero.*

*A **mis padres**, M^a de la Luz y José Manuel, por el apoyo recibido no solo en este, sino en todos mis proyectos profesionales y personales, por impulsarme a crecer, por mantenerse tan cerca en la distancia.*

*A **mis hermanos**, Teresa y José Manuel, por mantenerme viva, por hacerme reír en el más extremo cansancio, por recibirme siempre con un gigantesco abrazo y despedirse en todas las ocasiones con un te quiero.*

*A **mi tutor**, Carlos, por adaptarse a las diferentes circunstancias, por apoyar mis objetivos y enfocarme a cómo conseguirlos, por aceptar mis cambios, por ilustrarme en todas las ideas planteadas.*

*A **mis antiguos compañeros de trabajo**, el equipo de Gobierno y Control, por escucharme, por consolarme, por permitirme involucrarme por completo en este proyecto, por su sincera opinión, por valorarme.*

Sin ellos, este proyecto no hubiese sido posible

RESUMEN

El uso de técnicas de Big Data está ya extendido y madurado a nivel empresarial, es de conocimiento de todos que las grandes empresas se lucran utilizando este tipo de técnicas en sus análisis de datos para apoyarse en la toma de decisiones e intentar conocer mejor a su consumidor y su entorno.

En el presente estudio, planteamos la posibilidad de aplicar dichas técnicas en las pequeñas y medianas empresas. Para ello, haremos un recorrido por la historia del Big Data, además de definirlo y conocer las características que le diferencian del resto de estudios de análisis de datos convencionales. Todo ello nos permitirá entender cómo aplicarlo a las pequeñas y medianas empresas.

Para corroborar la información recogida durante todo el estudio, expondremos un ejemplo con las características de una empresa real, en el que utilizando herramientas o servicios que hacen uso de las técnicas de Big Data, mejoremos la toma de decisiones en dicha sociedad, reduciendo así costes para la empresa.

ABSTRACT

The use of Big Data techniques is already widespread and matured at the enterprise level, it is known that big companies obtain profit by using this type of techniques in their data analysis to support decision making and try to better understand their consumer and their environment.

In the present study, we propose the possibility of applying these techniques in small and medium enterprises. For this, we will take a tour of the history of Big Data, in addition to defining it and knowing the characteristics that differentiate it from the rest of studies of analysis of conventional data. All this will allow us to understand how to apply it to small and medium enterprises.

For corroborate the information collected in this study, we will present an example with the characteristics of a real company, where using tools or services that make use of Big Data techniques, we improve the decision making in this society, thus reducing costs for the company.

ÍNDICE

1. INTRODUCCIÓN	9
1.1. OBJETIVOS	9
1.2. JUSTIFICACIÓN DE LA NECESIDAD	9
2. BIG DATA	10
2.1. HISTORIA	10
2.2. DEFINICIÓN	13
2.3. HERRAMIENTAS	16
2.4. ÁMBITOS DE APLICACIÓN DEL BIG DATA	20
2.4.1. BUSINESS ANALYTICS	22
3. SMART DATA	23
4. BIG DATA EN PYMES	26
4.1. SMALL DATA	26
4.2. HERRAMIENTAS	30
4.3. COSTES DE IMPLANTACIÓN	33
5. EJEMPLO DEL USO DEL BIG DATA EN PYMES	34
6. CONCLUSIONES	36
6.1 CONCLUSIONES PERSONALES	38
7. BIBLIOGRAFÍA	39

ÍNDICE DE FIGURAS.

Figura 1. Previsión de tráfico de datos móviles.	11
Figura 2. Características del Big Data	15
Figura 3. Proyección de volumen de datos en empresas.	23
Figura 4. Relación existente entre Big Data y Smart Data.	25
Figura 5. Dispositivos electrónicos en autónomos y Pymes.	29

ÍNDICE DE TABLAS.

Tabla 1. Diferencias entre BBDD relacionales y BBDD no relacionales.....	17
Tabla 2. Diferencias entre Big Data y Small Data.	28

1. INTRODUCCIÓN

Iniciaremos el presente estudio exponiendo los objetivos que queremos conseguir con la realización del mismo, y justificando la necesidad de su realización. Es decir, expondremos a continuación el porqué es necesario realizar este Trabajo de Fin de Grado y qué aportará a las pequeñas y medianas empresas.

1.1. OBJETIVOS

En el presente Trabajo de Fin de Grado, estudiaremos la posibilidad de utilizar técnicas de análisis Big Data en pequeñas y medianas empresas.

Durante la lectura del trabajo, iremos comprendiendo qué son estas nuevas tecnologías, como podemos aplicarlas a las pymes y la necesidad de utilización real que existe. Los objetivos del presente estudio son demostrar la posibilidad de usar dichas tecnologías, además de herramientas o esfuerzos necesarios para poder beneficiarse. Pero, ¿Por qué queremos comprobar su uso en pymes?

Al igual que en las grandes empresas, las pymes necesitan aprovechar y obtener conocimiento de los datos que poseen. La toma de decisiones ha pasado de estar fundamentalmente basada en la experiencia para complementarse enormemente con el estudio de los datos recogidos de nuestros clientes, de procesos, o de productos y de servicios, por ejemplo.

Demostrando su uso y cómo pueden beneficiarse, las pequeñas y medianas empresas podrán conocer mejor su sector y sobre todo, su negocio.

1.2. JUSTIFICACIÓN DE LA NECESIDAD

“Lo que no se define no se puede medir. Lo que no se mide no se puede mejorar. Lo que no se mejora, se degrada siempre.” es una cita del físico y matemático británico William Thomson Kelvin, en la que pone de manifiesto la necesidad de medición y por consiguiente, de conocimiento.

Como hemos comentado anteriormente, en este estudio analizaremos si es posible que las pequeñas y medianas empresas utilicen, y por lo tanto, se beneficien de las nuevas tecnologías Big Data.

Ahora es posible gestionar nuestros riesgos; los riesgos asumidos se reducirán considerablemente, ya que podremos conocer, por ejemplo, si un cliente es o no solvente para aplazarle el próximo pago, si necesita un modelo específico de un televisor, o si le gustará nuestro próximo batido de fresa y canela. Es así como mediante la explotación de nuestros datos, conoceremos a nuestro cliente y sus futuras necesidades para satisfacerlas.

Este nuevo concepto del análisis de datos está consolidado en el mercado, y es de conocimiento de todos, que las grandes empresas se apoyan en dichas tecnologías para obtener nuevo conocimiento. Si las pequeñas y medianas empresas no comienzan a plantearse su posible uso, la brecha existente actualmente entre ambas, crecerá aún más.

Si podemos demostrar como la pequeña empresa hace uso del Big Data, tendrá la posibilidad de obtener una ventaja competitiva muy atractiva, puesto deja de basarse en intuiciones para comenzar a tomar decisiones basadas en datos.

2. BIG DATA

En este punto haremos un recorrido por la historia del análisis masivo de datos para así comprender el nacimiento, uso y características, que distinguen al Big Data del resto de estudios. Además, gracias a este apartado podremos comprender su concepto y hacernos una idea de la importancia que tiene su conocimiento y aplicación.

2.1 HISTORIA

Durante toda su historia, el hombre ha ido recabando datos e información para obtener un mayor conocimiento pero no es hasta el año 1997, cuando aparece por primera vez en un artículo en el que dos investigadores de la NASA, aseguraron que *“el ritmo de crecimiento de los datos empezaba a ser un problema para los sistemas informáticos actuales.”* A esto lo denominaron *“el problema del Big Data.”* (Cox y Ellsworth, 1997)

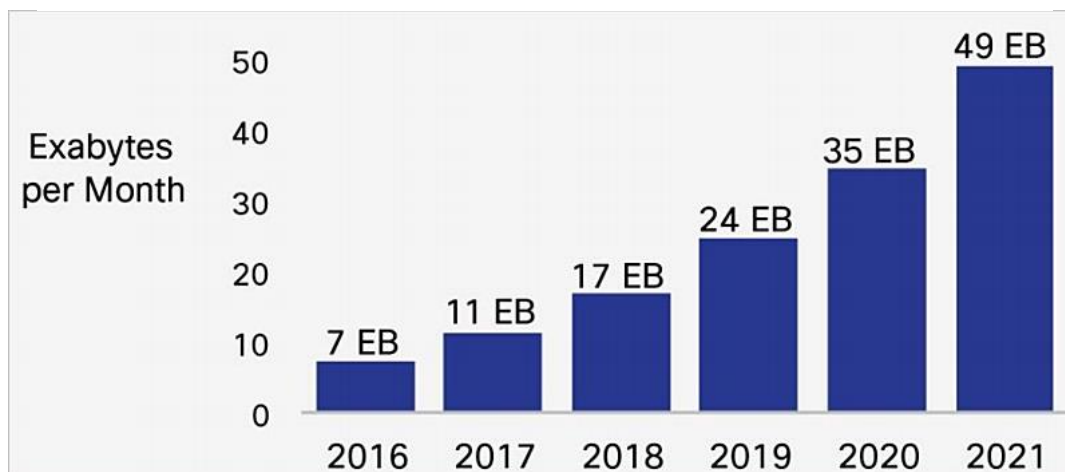
En un artículo titulado *The World’s Technological Capacity to Store, Communicate, and Compute Information* (Hilbert y López, 2011) se afirma que *“en el 1986 el 99,2% de los datos almacenados eran analógicos mientras que en 2007, el 94% del almacenamiento era digital”*. Con estos datos comprobamos el cambio radical que hemos sufrido en apenas 20 años.

Desde entonces, con el paso del tiempo, la cantidad de datos ha crecido de forma exponencial. El estudio *How Much Information?* afirma que “*la producción total de información en el mundo en el año 1999 fue de 1.5 billones de Gigabytes*” (Lyman y Varian, 2004).

El ritmo no ha parado de crecer ya que actualmente, según el Informe Cisco Visual Networking Index sobre el Tráfico Global de Datos Móviles 2016-2021, el pasado año, en 2016, “*se generó un tráfico de datos móviles de 7 exabyte al mes, lo que supone más de 84 exabyte al año*” (C. V. N. 2017)

Según el Informe y cómo se muestra en la siguiente imagen, se prevé llegar a los 49 exabyte al mes en 2021, lo que supondrá llegar a la mitad de un zettabyte anual.

Figura 1. Previsión de tráfico de datos móviles.



Fuente: C. V. N. 2017.

Estos datos son realmente sorprendentes y abrumadores, debido a que sólo nos estamos refiriendo al tráfico de datos móviles, que en 2021 supondrá tan solo el 20% del tráfico total de datos mundial. Es decir, tan sólo el tráfico de datos móviles, supondrá la mitad de un zettabyte en el año 2021.

Llegados a estos límites, ya no es tan importante el porcentaje de crecimiento y volumen de datos generados, puesto que entrados en tales dimensiones, no hay posibilidad de análisis y por

consiguiente, de extracción de información, sino lo hacemos mediante el uso de las nuevas tecnologías.

Y es la llegada de internet lo que actúa como catalizador del cambio. Es su democratización, su posible acceso desde empresas, organismos y hogares particulares, lo que hace explotar la generación continua de datos, dando comienzo a una nueva forma de vida.

La reducción de costes en la tecnología, la llegada de los Smartphones, de las redes sociales, de la geolocalización entre otros, y de que hoy en día ejercer un trabajo cualificado supone necesariamente trabajar o utilizar en algún momento del proceso un ordenador, es lo que ha supuesto el boom en la actual generación de datos.

Con este cambio cultural, nacen las nuevas empresas de servicios totalmente digitalizados, las cuales toda su actividad comercial se genera a través de internet, cuyo esquema de negocio se basa en dar acceso a sus usuarios a un gran número de recursos sin ser ellas las que los desarrollan, llegando así a poder vender en cualquier lugar del mundo sin necesidad de poseer un establecimiento físico. Aquí podemos poner de ejemplo empresas como puedan ser Amazon, Spotify, WhatsApp, Google, Netflix o Airbnb, que son líderes en sus respectivos sectores prestando sus servicios a los clientes través de internet.

Es así como generamos datos continuamente sin apenas percibirlo: pidiendo un taxi con el móvil, buscando los síntomas de la gripe, enviando un mensaje, escuchando una canción, o haciendo la lista de la compra compartida con un familiar mediante Google Drive, por ejemplo.

Es este vertiginoso aumento de información no estructurada, el detonante del uso de nuevas herramientas de almacenamiento y procesamiento de datos, que explicamos en el apartado 2.3. Dichas herramientas son las que hacen posible el análisis Big Data, permitiéndonos gestionar enormes cantidades de datos, desestructurados, a una velocidad nunca antes vista.

El análisis masivo de datos tal y como vamos a definir a continuación, se ha hecho posible en cualquier sector gracias a los grandes avances en la potencia de procesamiento, y la reducción de costes en la tecnología informática.

2.2 DEFINICIÓN

Cómo explicamos en el punto anterior, la tecnología y los diferentes usos de esta, están cambiando a un ritmo vertiginoso la forma de movernos, de comer, de comprar... En resumen, evolucionando rápidamente el método en el que resolvemos los problemas más cotidianos.

Es así poniéndonos en situación, como seremos capaces de entender el concepto de Big Data.

Según José Carlos López López director de operaciones de IMC Group, podemos definir el término de Big Data como *“La gestión y análisis de enormes volúmenes de datos que no pueden ser tratados de manera convencional, ya que superan los límites y capacidades de las herramientas de software habitualmente utilizadas para la captura, gestión y procesamiento de datos”*. (López, 2014).

Una gran diferencia es que en estadística, en los que estudios que se habían venido haciendo hasta ahora, se utilizan muestras estructuradas, mientras que en Big Data se examinan millones de datos a través de modelos o algoritmos cuyos resultados permiten una toma de decisiones mucho más precisa, con mayor conocimiento que el que podamos extraer de estudios tradicionales.

Algo que además caracteriza al análisis masivo es que, con frecuencia, estos datos se usan para un fin distinto del que en un inicio fueron recogidos y recopilados. Esto podemos explicarlo puesto que la gran revolución no está en la generación de datos, ni en el volumen, sino en la posibilidad de relacionar dichos datos entre sí para poder obtener una mayor información que la que en teoría se buscaba almacenar. Puede que en un principio recojamos y almacenemos la satisfacción de nuestros clientes para conocer si están contentos con el servicio, y puede que en un futuro apoyándonos en las nuevas tecnologías, utilicemos estos mismos datos junto con otra información para lanzar un nuevo producto, para abrir una nueva sucursal o para predecir la fuga de nuestros clientes.

A esto además, debemos añadir las características para que un conjunto de datos se considere tratado mediante técnicas Big Data. Como hemos definido anteriormente, llamamos análisis Big Data a un conjunto de datos que no pueda analizarse de forma convencional, no solo por la

cantidad de datos que generamos en la actualidad, sino además por sus ya conocidas V's (Douglas, 2001).

A continuación reflejaremos las 4 principales características que diferencian a un estudio en el que se han utilizado tecnologías Big Data, de un estudio en el que se han usado tecnologías convencionales: Volumen, Velocidad Variedad y Valor.

- El **Volumen** hace referencia a la cantidad de datos utilizados para realizar el análisis, quizás es la característica más comúnmente conocida, debido a que el primer problema que surgió fue el poder almacenar y analizar la cantidad de datos generada.

Para definir el volumen de análisis, un buen ejemplo es la creación de un modelo que prediga el impago (default) de un cliente de banca, cuanto mayor sea la cantidad de datos que poseemos, mejor podremos ajustar nuestro modelo, y por consiguiente, menor porcentaje de error en la clasificación del cliente.

- **Velocidad.** Esta característica refleja la rapidez con la que los datos son creados, almacenados y procesados a tiempo real. La ligereza con la que somos capaces de obtener información, nos ha permitido realizar multitud de análisis que de lo contrario, no aportarían ningún valor.

Un ejemplo de este atributo, es la posibilidad de enviar una oferta personalizada al móvil de nuestro cliente, cuando se encuentra a menos de 500 metros de nuestro comercio.

- Con la **Variedad** hacemos alusión a la diversidad de datos que podemos analizar en un mismo estudio; formas, tipos y fuentes en las que se pueden registrar los datos. Esta característica surgió como la necesidad de análisis de distintas fuentes.

Así por ejemplo, se nos permite usar en un mismo estudio datos de correos electrónicos, de encuestas, de documentos de texto, o incluso del audio de una llamada telefónica.

- El **Valor** es una de las principales características puesto que es el fin del estudio. Con las técnicas de análisis Big Data, podemos obtener nuevas relaciones entre los datos, lo que nos permite alcanzar nueva información que sin usar dichas técnicas, no podríamos conseguir.

El número de V's que se le atribuyen no está correctamente definido, ya que con el paso del tiempo y dependiendo de quien haga referencia a las características, se le atribuirán unas u otras.

En el siguiente gráfico hacemos referencia a 7 de las cualidades más importantes: Volumen, Velocidad, Variedad, Veracidad, Viabilidad, Valor y Visualización.

Figura 2. Características del Big Data



Fuente: Elaboración propia.

En este estudio solo explicamos y reflejamos las 4 primeras (Velocidad, Volumen, Variedad, Valor) puesto que son las que surgen en primera instancia y suponen un nuevo reto en el análisis de datos, mientras que el resto de atributos (Veracidad, Viabilidad y Visualización) ya han sido abordadas en los modelos de análisis convencionales. Valor se incluye en el primer grupo, no porque suponga un nuevo reto al estudio de los datos, sino porque gracias a las tecnologías Big

Data, somos capaces de obtener un valor oculto en la información, imposible de adquirir con otras herramientas.

Es cierto que aunque estas cualidades se pueden atribuir a otro tipo de estudios, hacemos referencia a ellas debido a que con la rapidez con la que pueden ser recopilados los datos, es importante comprobar su *Veracidad* para no conducir a errores en el futuro estudio. A la *Viabilidad* le hacemos mención para justificar el esfuerzo que pueda suponer a una empresa u organismo el utilizar las tecnologías Big Data, realmente los análisis con estas tecnologías aportan un valor aumentado a cualquier estudio, pero es importante demostrar que el estudio es viable para la empresa puesto que de lo contrario supondría un problema. Añadir *Visualización* se ha considerado significativo debido a que con la cantidad de datos que podemos poseer en un estudio, es revelador reflejarlos correctamente mediante tablas, gráficos u otros medios.

Así podemos concluir que, el análisis Big Data no es más que poder analizar y encontrar relaciones en grandes volúmenes de datos, que no poseen la misma estructura debido a que provienen de numerosas fuentes totalmente distintas, y a una velocidad que nos permite obtener valor de la información a tiempo real. Esta ha sido la verdadera revolución del análisis de datos, poseer mayor información, más rápidamente.

2.3 HERRAMIENTAS

Como hemos comentado anteriormente, el análisis masivo de datos ha sido posible gracias a las nuevas herramientas informáticas, al gran avance que hemos sufrido respecto al almacenamiento y procesamiento de datos, junto a la reducción de costes en la tecnología informática.

Entre las herramientas que hacen posible el uso y beneficio del Big Data, podemos destacar dos tipos; las herramientas de almacenamiento, y las de procesamiento de datos. En este apartado explicaremos en qué consisten y veremos algunos ejemplos de ellas.

Comenzaremos explicando las herramientas de almacenamiento en Big Data y las diferencias que encontramos con las BBDD tradicionales.

Habitualmente se han usado bases de datos como Oracle o MySQL, las cuales almacenan datos estructurados, lo que quiere decir, que los datos deben tener un formato establecido, como por ejemplo fechas o nombres. Las bases de datos no estructurados, surgen como necesidad de

almacenar datos originados de diversas fuentes, que no puedan almacenarse en tablas y que no deban tener necesariamente un formato concreto, como ocurre con los datos provenientes por ejemplo, de documentos de texto, de correos electrónicos o de archivos multimedia como un vídeo.

Entre las bases de datos no estructuradas o NoSQL (Not Only Sql) podemos destacar por ejemplo herramientas como Apache Cassandra¹, Apache Hbase² y BigTable³.

Podemos definir NoSQL como *“Una amplia clase de sistemas de gestión de bases de datos que difieren del modelo clásico de SGBDR (Sistema de Gestión de Bases de Datos Relacionales, en inglés, RDBMS) en aspectos importantes, siendo el más destacado que no usan SQL como lenguaje principal de consultas. Los datos almacenados no requieren estructuras fijas como tablas, normalmente no soportan operaciones JOIN ni garantizan completamente ACID (atomicidad, consistencia, aislamiento y durabilidad) y habitualmente escalan bien horizontalmente ”* (Lith y Mattsson 2010). Además podemos destacar que usan distintos modelos de datos, incluyendo los de gráficos, documentos, claves-valores y columnas.

En la siguiente tabla presentamos las diferencias existentes entre Bases de datos SQL (Relacional) y NoSQL (No relacional). (Amazon 2017)

Tabla 1. Diferencias entre BBDD relacionales y BBDD no relacionales.

	Bases de datos relacionales	Base de datos NoSQL
Modelo de datos	Normaliza los datos en tablas formadas por filas y columnas. Contienen un esquema que define estrictamente las tablas, columnas, índices, relaciones entre tablas y otros elementos.	Normalmente no contienen un esquema. Se suele utilizar una clave de partición para recuperar valores, conjuntos de columnas o documentos JSON o XML semiestructurados, así como otros documentos que contengan atributos de elementos relacionados.

¹ Página oficial de Cassandra: <http://cassandra.apache.org/>

² Página oficial de Hbase: <https://hbase.apache.org/>

³ Página oficial de BigTable: <https://cloud.google.com/bigtable/>

	Bases de datos relacionales	Base de datos NoSQL
Propiedades ACID	Admiten un conjunto de propiedades definidas por el acrónimo ACID: atomicidad, consistencia, aislamiento y durabilidad.	Se intercambian algunas de las propiedades ACID por un modelo de datos más flexible que se escala de forma horizontal.
Desempeño	El desempeño depende del subsistema de disco. Es necesaria la optimización de consultas, índices y estructura de tabla para lograr el máximo desempeño.	El desempeño es, por lo general, depende del tamaño del clúster de hardware subyacente, la latencia de red y la aplicación que efectúa la llamada.
Escalado	Lo más sencillo es ampliar la escala con un hardware más rápido. Se requieren inversiones adicionales para que las tablas relacionales abarquen un sistema distribuido.	Está diseñada para reducir la escala utilizando clústeres distribuidos de hardware de bajo costo para aumentar el desempeño sin que aumente la latencia.
API	Solicita almacenar y recuperar datos que están comunicados mediante consultas que se ajustan a un lenguaje de consulta estructurado (SQL por sus siglas en inglés). Estas consultas son analizadas y ejecutadas por los sistemas de administración de bases de datos relacionales (RDBMS).	Las API basadas en objetos permiten a los desarrolladores almacenar y recuperar fácilmente estructuras de datos en memoria. Las claves de partición permiten que las aplicaciones busquen pares de clave-valor, conjuntos de columnas o documentos semiestructurados que contengan atributos y objetos de aplicación serializados.
Herramientas	Normalmente ofrecen un amplio conjunto de herramientas que simplifican el desarrollo de aplicaciones de base de datos.	Ofrecen herramientas para administrar los clústeres y el escalado. Las aplicaciones representan la interfaz primaria de los datos subyacentes.

Fuente: Amazon, 2017.

La principal característica que nos encontramos en las herramientas de Big Data, es que son sistemas distribuidos. Cuando hablamos de distribución en los sistemas informáticos, hacemos

referencia a un conjunto de máquinas que están conectadas a través de una red de comunicaciones, por lo que se puede dividir la carga de procesamiento entre ellas. (Tanenbaum y Van Steen, 2007)

Esto supone que para poder procesar terabytes de información, no necesitamos una computadora gigante que sea capaz de soportar tal volumen, sino que distribuimos la información entre los distintos nodos que conforman el cluster. El que exista herramientas distribuidas nos permite almacenar o procesar una cantidad de datos que hasta ahora, solo era posible a unos grandes costos

A continuación definiremos dos de las principales herramientas existentes en el procesamiento de datos en Big Data.

Apache Hadoop (Hadoop Wiki, 2016) y Apache Spark (Zaharia, Chowdhury, Franklin, Shenker y Stoica, 2010) son herramientas distribuidas, con escalabilidad horizontal y tolerante a fallos.

Cuando hablamos de escalabilidad horizontal nos referimos a que el modo de obtener mayor capacidad de cómputo, es añadiendo nodos al cluster, al contrario que en escalabilidad vertical, que obtenemos mayor capacidad de computación aumentando la potencia de una misma máquina. Que las herramientas sean tolerantes a fallos, significa que si por algún motivo alguno de los nodos sufriera un fallo, seguiríamos trabajando con los nodos restantes sin grandes repercusiones.

Apache Hadoop fue el primero en surgir con este tipo de características, seguido de Spark. Ambos se diferencian principalmente en que Hadoop utiliza de manera intensiva lectura y escritura a disco, es decir, utiliza el disco de los nodos para guardar los pasos intermedios. En cambio, Spark utiliza la memoria RAM, por lo que esto puede suponer que Spark en ocasiones supere hasta 100 veces el rendimiento de Hadoop.

Las herramientas comentadas en este punto, no son más que algunos ejemplos de las que han ido surgiendo en los últimos años. Así, se refleja la revolución que hemos sufrido en la capacidad de almacenamiento y de cómputo. Es gracias a este tipo de software, que actualmente somos capaces de analizar grandes volúmenes de datos, rápidamente y desde distintas fuentes.

2.4 ÁMBITOS DE APLICACIÓN DEL BIG DATA.

Una vez definido el concepto, en este apartado, veremos hasta donde se han beneficiado los diferentes sectores de su uso.

Como hemos comentado anteriormente, gracias a las herramientas informáticas y capacidad de procesamiento, las tecnologías Big Data se han extendido a todos los ámbitos. Su uso se ha dilatado con tal rapidez, que no solo es recomendado sino que empieza a ser necesario.

Actualmente, las tecnologías Big Data son totalmente maduras, ya no se consideran tecnologías emergentes ni disruptivas sino que están totalmente consolidadas. *“El reto ahora es que el Big Data sea maduro a nivel empresarial”*, comenta Artigas (2016), quién además ha señalado que *“un 97% de los proyectos de Big Data están liderados por el CEO y solo un 3% los encabezan las áreas de IT de las empresas”*. (Artigas, 2016)

A continuación, veremos algunos ejemplos de uso en distintos sectores, ejemplos concretos de mejora, de valor adquirido en el sector, gracias al uso de estas técnicas:

- **Seguridad Ciudadana.** Cada vez más se recogen y analizan datos de los ciudadanos para prestar un mejor servicio. Un buen ejemplo en este ámbito, es Precobs, (Spiegel, 2014) un software que ha sido probado en las ciudades de Múnich y Nuremberg. Precobs recoge y analiza datos de crímenes para después, utilizando técnicas de Big Data, encontrar patrones en ellos, de manera que puede predecir cuándo y dónde se producirá el próximo delito para así desplegar un equipo y poder evitarlo.
- **Educación:** En AltSchool⁴ están siendo los pioneros en educación. Está formado por 4 escuelas de San Francisco que actualmente se dedican a recopilar datos de sus estudiantes como por ejemplo, la asistencia, en que invierten el tiempo que poseen para trabajar de forma independiente en temas que previamente ha fijado el profesor, o la evolución personal de cada estudiante.

⁴ Página oficial de AltSchool: <https://www.altschool.com/>

Actualmente está en progreso, pero su fin es crear un sistema que permita al profesor diseñar clases personalizadas y por lo tanto más efectivas, mediante el estudio de sus alumnos y con técnicas de análisis de Big Data.

Dentro de poco, el que todos los alumnos aprendiesen lo mismo al mismo ritmo, nos parecerá algo descabellado.

- **Cine:** El cine ya está usando estas tecnologías, y es un perfecto ejemplo de conocimiento del cliente y de sus necesidades o deseos. Algunos ejemplos de esto son; cómo la serie “Juego de Tronos” anticipó su éxito analizando las descargas ilegales, cómo “House of Cards” hizo un cambio de guion cuando la serie ya estaba emitiéndose, basándose en la información recogida de los espectadores que dieron un vuelco a la serie o cómo “Narcos” se produjo gracias a que Netflix cuenta con *“un algoritmo de recomendación capaz de saber cuánto nos va a gustar una serie o película antes de que la veamos, pueden aprobar guiones o hacer el casting de los actores en base a ello”* (Pérez, 2017). *“La lista de puntos de medida es enorme y cada vez mayor. Pueden inferir incluso si estamos en el camino de vuelta a casa, cenando o viendo una serie sólo para quedarnos dormidos. Toda esta información se utiliza para clasificar los contenidos por usos y perfiles”* (Pérez, 2017). Estos programas de éxito muestran un claro ejemplo del beneficio adquirido en el sector del cine usando las tecnologías Big Data.

- **Sanidad:** Las aplicaciones en este sector pueden ser muy amplias debido a que es el que mayor cantidad de datos genera a diario. Un ejemplo de aplicación de Big Data en sanidad es Savana.⁵

Savana es un software que con técnicas Big Data, lee numerosas historias clínicas, encontrando relaciones y dando así un servicio de salud personalizado y predictivo de enfermedades. Según Hernández (2017), neurólogo del Ramón y Cajal y fundador de Savana, *“Hoy en día tenemos computación, redes y sistemas, que nos permiten agregar y explotar nueva información para enfrentarnos a nuevos casos, investigar y gestionar mejor los hospitales.”* añade además que *“la medicina individualizada no es más que el aterrizaje del Big data en Sanidad y además de manera predictiva.”* (Hernández, 2017)

⁵ Página oficial de Savana: <https://www.savanamed.com/>

2.4.1 BUSINESS ANALYTICS

El sector que actualmente mayor beneficio obtiene con las nuevas tecnologías es el sector empresarial. En este apartado introducimos el Business Analytics (BA) puesto que supone una parte importante de la aplicación del Big Data en la empresa.

El concepto Business Analytics es una evolución del término Business Intelligence o Inteligencia de Negocio. Dresner (1993) definió « Business Intelligence (BI)» como *“los conceptos y métodos que mejoran la toma de decisiones de negocio mediante el uso de sistemas de apoyo basados en datos reales”*.

En 2009 Michael J. Beller, en su publicación ‘Next Generation Business Analytics’ explica el Business Analytics como *“los conocimientos, tecnologías y prácticas para la investigación y exploración continuamente interactiva del rendimiento del negocio, para ganar visión y capacidad de dirección en la planificación del negocio”*. (Beller y Barnett, 2009)

Podemos deducir de la definición anterior que es por tanto una nueva visión del negocio que no solo se basa en datos, sino también en métodos estadísticos para medir el pasado y poder conducir el futuro. La principal diferencia que encontramos con el BI (Business Intelligence) es que este último, tradicionalmente se focaliza en el uso y seguimiento de métricas o KPI’s.

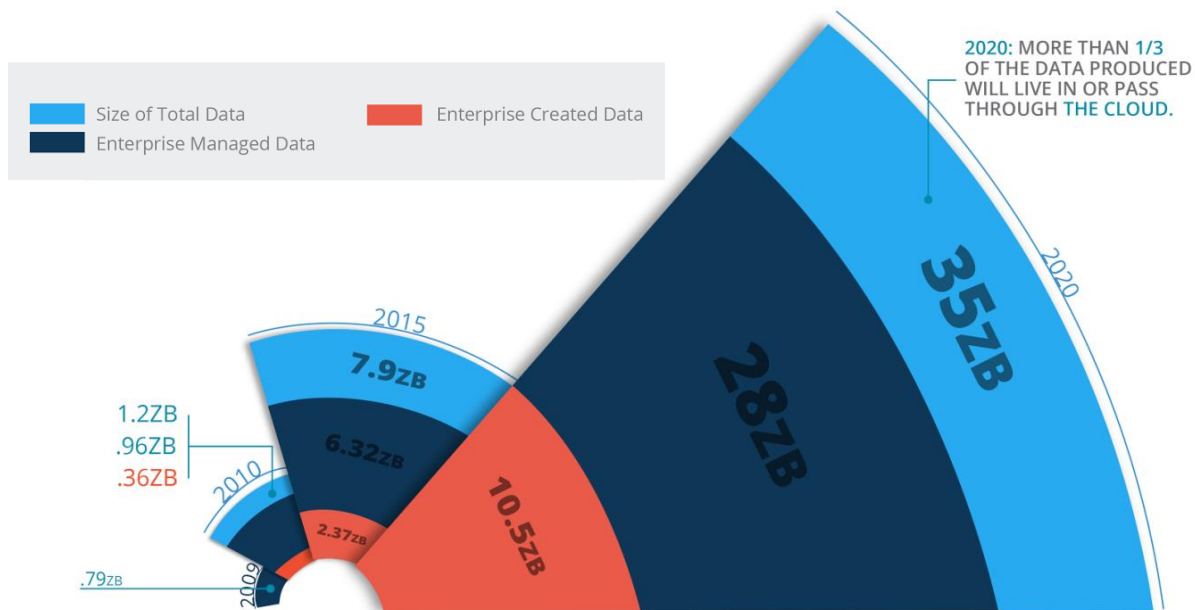
Es la unión del Big Data con el BA (Business Analytics) lo que genera un auténtico valor a la empresa. Si somos capaces de utilizar test estadísticos, de encontrar correlaciones en los datos, visualizar patrones o de crear modelos predictivos, unido a las características que nos presentan los análisis de Big Data, estaremos obteniendo un enorme valor de la información a tiempo real, lo que nos permitirá conocer a nuestros consumidores, y tomar las decisiones basándonos en ellos.

En el año 2009, la Inteligencia empresarial pasó a ser una de las principales prioridades para los directores de tecnologías de la información (Gartner, 2009) y actualmente prácticamente todos los sectores se lucran con ella: compañías de seguros, banca, marketing, consultoría, agencias de publicidad...

Como se muestra en el siguiente gráfico, los datos generados por empresas han aumentado enormemente en los últimos años, por lo que hoy en día es totalmente necesario gestionarlos y

explotarlos para obtener información de nuestro negocio, conocer el sector y las necesidades de los clientes potenciales.

Figura 3 Proyección de volumen de datos en empresas .



Fuente: DXC.technology.

En este sector podemos exponer numerosos ejemplos de uso de las nuevas tecnologías obteniendo beneficios para la empresa, pero en el capítulo 5 del presente estudio, pondremos de manifiesto algún ejemplo de como el uso del BA junto con el Big Data, se aplica a pequeñas y medianas empresas.

3. SMART DATA

En este apartado explicaremos el Smart Data, nuevo concepto que surge a partir del Big Data, qué es lo que diferencia a ambos términos y como se ha llegado a utilizar dicha idea.

El Smart Data es una ‘evolución’ del Big Data, en la cual, en este nuevo concepto se hace especial referencia a una de las v’s, al valor. Como hemos comentado anteriormente, la gran

revolución surgió de la posibilidad de poder realizar nuevas exploraciones en los datos y encontrar nuevos patrones o relaciones entre ellos, más que en el volumen que poseamos.

El concepto de Smart Data hace referencia a *“La transformación de largas listas de números y datos, en información con valor, con respuesta útil. Convertir datos en estadísticas utilizando complicados algoritmos que procesan el Big data.”* (Cegos, 2017)

La línea que separa ambos conceptos no está totalmente definida, ya que en ocasiones se pueden confundir. Podemos afirmar que el Smart Data pertenece y complementa al Big Data, es un término que comienza a cobrar mucha importancia y sin el cual el Big Data no tendría utilidad.

La diferencia podemos marcarla en que el Big Data se originó con las nuevas herramientas de almacenamientos y procesamiento, fue el boom que suponía poder gestionar un volumen desorbitado de datos, nació como la resolución al problema de tanta información, como la posibilidad de explotar grandes cantidades de información. El Smart nace posteriormente, cuando ya conocemos como almacenar, procesar y visualizar tal cantidad de información, tenemos resuelto el problema, por lo que no le damos tanta importancia al volumen que poseemos, si no al conocimiento que somos capaces de obtener con su estudio.

Con la llegada del análisis masivo, muchas empresas comenzaron a “ahogarse” en datos, cómo las nuevas tecnologías permiten su almacenamiento, se comenzó a recabar todo tipo de información, sin tener en cuenta cual deberían ser utilizada para el posterior estudio, sin preguntarse que se quería conseguir y sin saber cuáles deberían ser utilizados posteriormente.

Es por este motivo, por el que se comienza a hablar del Smart Data, que hace referencia a que no es necesario utilizar todos los datos, que no debemos centrarnos en recopilar la mayor cantidad de información posible sino, que lo que necesitamos es conocer en primer lugar, que problema queremos resolver o que nueva información queremos obtener, y así decidir qué datos necesitaremos almacenar y procesar para nuestro estudio.

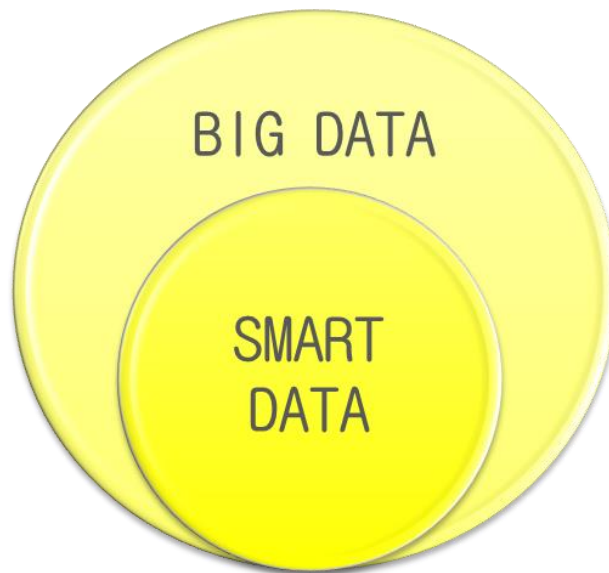
Con esta evolución del concepto, pasamos de “los grandes datos” a “los datos inteligentes”. Este nuevo término quiere transmitir que no necesariamente debemos almacenar y procesar Petabytes de información, podemos utilizar las nuevas tecnologías con un menor volumen, con el único fin de adquirir conocimiento.

No son conceptos ni mucho menos excluyentes, el deseo del análisis Big Data es utilizar los datos de manera inteligente y por consiguiente utilizar a su vez el Smart Data.

Para poder hacer estudios Smart, por supuesto necesitamos las nuevas tecnologías, las características del Big siguen estando presentes, puesto que las fuentes pueden seguir siendo varias, sin formatos comunes, y la velocidad de almacenamiento y procesamiento no será la que se ha venido usando en estudios convencionales.

Como mostramos en la siguiente imagen y como hemos comentado durante este apartado, el Smart Data, forma parte del Big Data. Es el fin de realizar estudios con dichas tecnologías, porque realizarlos sin pensar en primera instancia y sin priorizar el valor que vamos a obtener, supondría obtener análisis ineficaces.

Figura 4. Relación existente entre Big Data y Smart Data.



Fuente: Elaboración Propia.

Podemos concluir que las nuevas tecnologías Big Data son el conjunto del nuevo paradigma de análisis de datos, mientras que el Smart Data utiliza dichas tecnologías sin tener en cuenta el volumen de datos, centrándose sobre todo en el problema a resolver y utilizando los datos que realmente aportan valor a nuestro estudio.

4. BIG DATA EN PYMES

En este punto del estudio, revisaremos el concepto del Big Data, pero en este caso aplicado a pymes.

Reflejaremos cómo se definen las nuevas tecnologías en las pequeñas y medianas empresas, si las herramientas usadas para los estudios de análisis son las mismas, o en caso de no serlo, cuáles son las diferencias que presentan las herramientas en ambos términos. Veremos además, cuál es el balance de usar Big Data en este tipo de negocios, hablaremos de los costes que les suponen y los beneficios que puedan obtener.

En definitiva, si para las pequeñas y medianas empresas usar las nuevas herramientas de análisis es rentable y atractivo.

4.1. SMALL DATA

El desconocimiento de las técnicas y beneficios que ofrece el Big Data junto con su gran inversión, son los principales motivos por los que las pequeñas y medianas empresas no apuestan por estas nuevas tecnologías.

Así un estudio realizado por GTAI (2016) indica que: *“El 63% de las pequeñas y medianas empresas desconocen cuáles son las aplicaciones que supone el Big Data, mientras que el 10% considera que este recurso es irrelevante para su negocio. Sólo el 26% de las Pymes asegura entender los beneficios que tiene adoptar este tipo de herramientas en sus modelos de trabajo.”*

Las pymes pueden verse sobrepasadas por el Big Data, pensar que no disponen de los recursos necesarios para llevarlo a cabo, o que los esfuerzos necesarios en caso de obtener dichos recursos, no compensan con el beneficio obtenido. Es bajo este entorno vamos a definir el Small Data, que podemos adelantar como el “Big Data en pymes”.

En el artículo *“Small Data: pequeñas pistas que aportan mucho valor a tu negocio”*, Josep Martínez Vicedo, director de marketing de AINIA, centro tecnológico, define el Small Data como: *“Un concepto que se puede definir como el proceso por el cual hacemos uso y análisis de datos de menor tamaño y simplicidad (a los de Big Data), que afectan a las particularidades de nuestro negocio, que forman parte de la rutina diaria y permiten conocer mejor lo que sucede en*

nuestro entorno así como identificar oportunidades para la eficiencia de los procesos y la relación con los clientes.” (Martínez, 2016).

En el capítulo 3 del presente estudio, hemos definido un nuevo concepto, Smart Data, en el que lo detallábamos como el uso de técnicas de Big Data poniendo el valor logrado del análisis en el centro, como único fin, sin importar la cantidad de datos que tengamos para realizarlo. Partiendo de esta base, podemos definir el Small Data como el uso de técnicas de análisis de Big Data en pequeñas y medianas empresas.

Al igual que en el Smart Data, aquí se pone de manifiesto que no es necesario almacenar y procesar millones de gigabytes de información para llegar a encontrar patrones en nuestros datos y conocer a nuestros clientes o negocio.

En el Smart, el fin único es el conocimiento, priorizar que información queremos adquirir antes de comenzar el análisis, sin tener en cuenta el volumen de datos utilizado. El Big supone analizar una cantidad abrumadora de datos con las herramientas que nos permiten procesar tal cantidad de información conjuntamente, para encontrar patrones entre ellos, llegar a obtener información que un principio no creíamos que fuese el fin del estudio. El Small es una combinación de ambos, es usar herramientas de Big Data a una menor escala, con un volumen bastante inferior, focalizarnos en nuestro objetivo y conocer que datos necesitaremos para realizar el análisis puesto que recoger y almacenar todo tipo de datos puede suponernos un esfuerzo que no llegue a verse reflejado en los beneficios.

En una entrevista realizada a Martin Lindstrom, escritor sobre el comportamiento del consumidor y escritor del libro “Small Data: Las pequeñas pistas que nos advierten de las grandes tendencias” asegura que el “*Big Data consiste en encontrar correlaciones, mientras que el Small Data se preocupa por las causas, las razones detrás de las cosas*” (Lindstrom, 2016). Durante su entrevista pone de manifiesto que tan importante es utilizar grandes volúmenes de datos y encontrar correlaciones entre ellos que nos den información sobre nuestro negocio, cómo seguir pequeñas pistas que nos acerquen al consumidor. Lindstrom (2016) Afirma además que “*en ocasiones, algunas empresas se ven sobrepasadas por el Big Data, que no todo consiste en analizar millones de datos con algoritmos*”, y en esto, la pyme juega con ventaja.

Las características que definen el Big Data, (las V's) aplican también al Small Data a diferencia de la primera, el Volumen. En el Small Data el volumen es menor puesto que las pequeñas y medianas en empresas no generan la misma cantidad de datos que las grandes organizaciones. El resto de atributos son deseables y aplicables en los análisis con un volumen menor de datos.

En el siguiente cuadro mostramos las diferencias entre el Big y Small Data.

Tabla 2. Diferencias entre Big Data y Small Data.

	Big Data	Small Data
Fuentes de datos	Datos generados fuera de la empresa a partir de fuentes de datos tradicionales: - Datos de Registro - Datos de comunicación social - Datos de dispositivos - Vídeos, imágenes, etc.	Datos empresariales tradicionales: - Recursos empresariales. - Planificación transaccional de datos - Relación con el cliente - Datos financieros - Transacciones web
Volumen	- Terabytes - Petabytes - Exabytes - Zettabytes	- Gigabytes - Terabytes
Velocidad	- A menudo a tiempo real - Requiere respuesta inmediata	- Batch o próximo a tiempo real - No siempre se requiere una respuesta inmediata
Variedad	- Datos Estructurados - Datos No Estructurados - Datos Multi-estructurados	- Datos Estructurados - Datos No Estructurados
Valor	- Complejo, avanzado, predictivo, business analysis.	- Business intelligence, analysis y reportes.

Fuente: Banafa, 2014

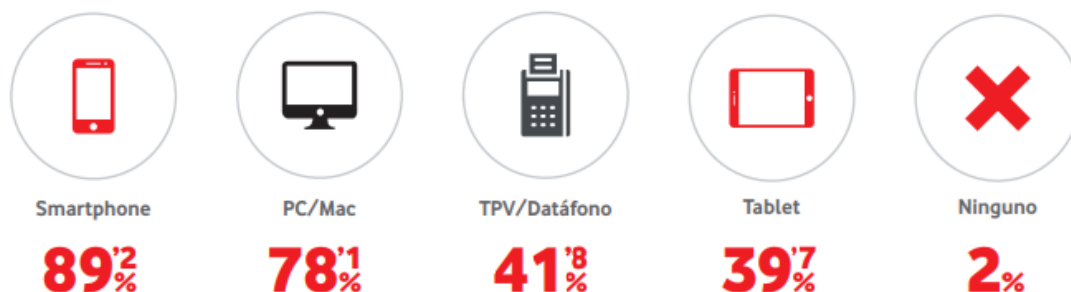
En el anterior cuadro, se presentan las diferencias entre ambos análisis, que como podemos observar las principales aparecen en el volumen y en las fuentes de datos. En Big Data, es deseable un volumen de datos, lo que implica analizar datos de fuentes tanto externas como

internas a la empresa, mientras que en el Small Data no se tiene tan en cuenta en volumen por lo que se normalmente proceden de fuentes internas.

El resto de características presentan diferencias pero no son significativas, puesto que por ejemplo, la velocidad en Small data no es a tiempo real pero se aproxima mucho, se puede analizar mediante Batch debido a que el flujo de información no es tan continuo como en el Big Data. El valor adquirido del estudio no es tan complejo cuando hablamos de Small Data, debido a que como hemos comentado anteriormente, es conveniente fijar los objetivos a conseguir, por ello, analizar los resultados del análisis nos resultarán más sencillo que si no se ha fijado previamente.

Las pequeñas y medianas empresas españolas se están digitalizando, se apoyan en servicios o productos digitales para gestionar de manera más eficiente su negocio, lo que facilita enormemente la recogida de datos para su posterior análisis.

Figura 5. Dispositivos electrónicos en autónomos y Pymes.



Fuente: Vodafone 2016

La imagen anterior aparece en un estudio realizado por porquetienesqueestar.com, El Observatorio Vodafone de la empresa. En la imagen se muestra como tan solo un 2% de las pymes españolas no utiliza ninguno de los dispositivos indicados en la imagen. Casi un 90% de las empresas usa el Smartphone y aproximadamente un 80% utiliza pc para facilitar el trabajo diario.

Lo datos mostrados anteriormente nos indican que es posible llevar a prácticamente todos los negocios el Big o Small Data, ya que no es de gran dificultad, utilizando este tipo de dispositivos, recopilar los datos que sean necesarios para nuestro futuro análisis.

El obstáculo principal que tienen las pymes a la hora de incorporar técnicas y análisis Big Data en sus procesos, es la gran inversión necesaria para ello. Por eso, es de vital importancia detectar los focos de mejora en la organización, hacer la pregunta correcta y conocer la información necesaria para conseguir los objetivos. Por ejemplo, definir si se quiere seguir una nueva estrategia, si lo que se busca es identificar un nuevo servicio/producto el cual los clientes aceptarán, o queremos conocer su satisfacción.

Identificar el objetivo a conseguir, permitirá identificar la información necesaria para llevarlo a cabo. Conocer de qué información disponemos y de cual carecemos, nos permitirá identificar como hacer uso del Big Data y cuándo podremos ponerlo en marcha.

La cercanía que tiene la pyme con su cliente, junto con los estudios de datos que actualmente tenemos la posibilidad de hacer, presentan una verdadera ventaja para el negocio. Si unimos ambas fortalezas, podemos llegar a conocer muy ampliamente a nuestro consumidor y sus futuras necesidades.

El objetivo ahora para poder emplear técnicas de Big Data en pymes, es conseguir que los datos obtenidos de distintas fuentes, se homogeneicen y trabajen de conjuntamente para llegar a un fin común.

4.2. HERRAMIENTAS

El gran reto para que las pymes puedan beneficiarse de las técnicas de análisis de Big Data, está en conocer las distintas herramientas y métodos que más se ajusten a su negocio, el potencial de sus empleados y sobre todo, sus datos y el tipo de estudio que quieran realizar.

Antes de comenzar a buscar una solución, debemos preguntarnos cuál es el problema. En primer lugar debemos tener muy claro cuáles son los objetivos a conseguir analizando nuestros datos y utilizando este tipo de herramientas, seguidamente debemos pensar que compartir nuestros datos

con otras empresas para poder analizarlos, no nos perjudica, no es una vía de escape de información, todo lo contrario. Tener presentes nuestras fortalezas y debilidades en la explotación de la información nos permitirá encontrar la mejor solución adaptada a nuestro negocio, pedir asesoramiento si lo necesitamos, proporcionará la visualización de nuestros datos.

A continuación detallamos algunas herramientas que ayudan a la gestión de datos de la pequeña y mediana empresa, permiten conocer a nuestros clientes o su comportamiento sin necesidad de grandes conocimientos técnicos.

- **Google Analytics**⁶. Esta herramienta es de fácil uso, se trata de analizar los movimientos que han surgido en nuestra página web. Es interesante porque nos da datos como de donde provienen las visitas, el tiempo que han permanecido en la página, si finalmente realizaron o no una compra, entre otros datos. Conociendo como está funcionando nuestra página web y si a través de ella se están realizando o no compras, podemos aplicar distintas estrategias de marketing para mantener esos datos o para darle un impulso nuestra imagen en la web.
- **Alexa**⁷. Alexa, también es una herramienta enfocada a mejorar nuestra página web. Pertenece a Amazon y nos ayuda por ejemplo a mejorar nuestro SEO, nos proporciona información sobre ideas para hacer crecer nuestro tráfico, encontrar palabras clave de fácil rango o comparar nuestra página web con la de nuestros competidores, entre otras funciones. Alexa es una herramienta ideal para mejorar nuestro posicionamiento web o compararlo con la competencia.
- **SurveyMonkey**⁸. Esta herramienta nos permite hacer encuestas a través de correo electrónico, es intuitiva y fácil de usar. Dependiendo de dónde se quiera llegar, serán necesarios más o menos conocimientos sobre la realización de encuestas y análisis de datos. Esta herramienta nos permite visualizar los resultados de la encuesta que hemos lanzado a tiempo real. Presenta gráficos fáciles e intuitivos que se actualizan cada vez que

⁶ Página oficial Google Analytics: <https://www.google.es/intl/es/analytics/>

⁷ Página oficial Alexa: <http://www.alexa.com/>

⁸ Página oficial Survey Monkey: https://es.surveymonkey.com/mp/take-a-tour/?ut_source=header

se añade una respuesta, además se puede comprobar quién no ha respondido aún y enviar recordatorios. SurveyMonkey es una buena herramienta para realizar preguntas a nuestros clientes y obtener la información a tiempo real.

- **Commerce 360⁹ o Mi comercio¹⁰** del BBVA y Banco Santander respectivamente, analizan información recogida del TPV de la empresa, de manera que tenemos un conocimiento muy amplio sobre nuestros clientes. Estas herramientas nos ofrecen dossieres con información como por ejemplo: la fidelidad de los clientes, frecuencia de compra en nuestro establecimiento, segmentos demográficos, comparaciones con las ventas en la competencia de la zona, entre otros servicios, que nos permitirán seguir la estrategia que mejor se adapte a nuestro negocio. Además, debemos añadir que ambas herramientas son gratuitas para los clientes del respectivo banco, por lo que acceder a dicha información es realmente fácil.
- **Nebinn¹¹** se caracteriza por presentar “soluciones Big Data al alcance de todos”. Es una herramienta que permite explorar y visualizar nuestra información recopilada de distintas fuentes y a tiempo real. Además de visualizar nuestros datos, aplica modelos predictivos sobre la información de nuestro negocio, que nos facilitará la toma de decisiones y el conocimiento sobre nuestros clientes. Esta herramienta es muy atractiva por el tipo de información que nos permite conocer, además de por ser fácil e intuitiva, consta de un equipo de expertos en Big Data que hace posible el análisis de nuestros datos a tiempo real y de fuentes heterogéneas.

Además existen otro tipo de herramientas más especializadas y que necesitan de mayor conocimiento técnico como por ejemplo:

- **Pentaho¹² y Jaspersoft BI¹³**. Estas herramientas están especializadas en la integración de datos recopilados de distintas fuentes. Es de gran ayuda cuando tenemos que consolidar y

⁹ Página Commerce 360: <https://www.commerce360.es/>

¹⁰ Página Mi Comercio: <https://www.bancosantander.es/es/empresas-advance/banca-online/app/mi-comercio>

¹¹ Página oficial de Nebinn: <http://nebinn.com/>

¹² Página oficial de Pentaho: <http://www.pentaho.com/>

¹³ Página oficial de Jaspersoft: <https://www.jaspersoft.com/>

analizar toda la información en una sola plataforma. Además destacan por contar ya con versiones open source.

En este apartado hemos puesto de manifiesto que antes de comenzar a utilizar ninguna herramienta, debemos conocer en profundidad que beneficio queremos conseguir o que problema queremos solucionar. Además, hemos mostrado algunos ejemplos de herramientas que existen en el mercado y que utilizan técnicas Big Data. Como hemos visto durante el desarrollo de la presente sección, existen soluciones muy amplias para el análisis de datos en la pequeña y mediana empresa.

4.3. COSTES DE IMPLANTACIÓN

“La caída de los precios en banda ancha y almacenamiento de datos permite a las empresas, por un lado, invertir en Big Data y, por otro, la democratización en el procesamiento de datos. En el año 2000, abrir una start up en Silicon Valley requería 5 millones de dólares. Hoy esta operación cuesta 5.000 dólares: la razón es el uso de software libre” (Artigas, 2016)

Con el párrafo anterior queremos poner de manifiesto cómo se está llevando a cabo una democratización real en el procesamiento de datos. Los costes de almacenamiento y procesamiento de datos son cada vez menores, por lo que esto nos permite obtener un gran valor a un coste cada vez más bajo.

Esto supone una gran oportunidad para la pequeña y mediana empresa, ya que no tiene los mismos recursos que las grandes para invertir en formación del personal y herramientas que hagan posible la explotación de los datos.

Actualmente existen numerosos cursos para formarnos en este ámbito, algunos podemos encontrarlos gratuitos, como es el caso por ejemplo, de los cursos de la plataforma Actívate de Google¹⁴. En dicha plataforma podemos encontrar cursos sobre competencias digitales, algunos de ellos son por ejemplo: “Curso online de Analítica Web”, “Curso online de Cloud Computing” o “Curso de Marketing Digital”. Esto nos permite ampliar nuestros conocimientos o el de

¹⁴ Página oficial de la plataforma Actívate: <https://www.google.es/landing/activate/formate/>

nuestros empleados, para comenzar a sumergirnos en competencias digitales y acercarnos al análisis de datos mientras técnicas de Big Data.

En cuanto a las herramientas, como hemos visto en la sección anterior, algunas son open source, y muchas de las mencionadas anteriormente tienen distintas versiones a diferentes precios que se ajustan a nuestras necesidades. Además de las herramientas reflejadas en el presente estudio, existe una gama bastante amplia a unos precios muy competitivos adaptados a las necesidades de cada negocio.

En conclusión a este apartado, queremos reflejar que existen distintas alternativas para las pymes, que se adaptan a sus exigencias y que pueden elevar de manera considerable sus beneficios gracias a los nuevos conocimientos adquiridos con los estudios realizados.

5. EJEMPLO DEL USO DEL BIG DATA EN PYMES

En esta sección del estudio expondremos un ejemplo de cómo utilizando algunas de las herramientas explicadas en el apartado anterior, podemos aplicar técnicas de Big Data en pymes.

El escenario, será una cadena de 5 cafeterías con 45 empleados en total y con un balance de negocio de aproximadamente 700.000€ anuales entre todas las cafeterías.

Cómo se ha comentado durante la realización del presente estudio, el paso inicial es conocer de qué carecemos y con qué recursos contamos, para poder adquirir nueva información.

En este caso, se aprobó unas tarjetas de descuento asociadas a cada uno de los clientes, para que cada vez que consuman en alguna de las cafeterías, la muestren y se les aplique un 5% de descuento en el importe total de ese día. De este modo, se tiene identificados datos del cliente como por ejemplo; fecha y hora de la consumición, qué productos consumen en cada visita o si fueron solos o acompañados por más personas. Además, debemos añadir que esta cadena de cafeterías, pasó de tomar las comandas de las consumiciones de manera manual, a utilizar tablets para anotar los productos que los clientes querían tomar en ese momento. Así, además de tener almacenados que productos se han consumido por día y hora, podemos conocer los tiempos de espera o afluencia de clientes, entre otros.

Añadimos que se tiene completamente digitalizado mediante el uso de herramientas office, las compras a proveedores; frecuencias de compra, costes, entrega de la mercancía, entre otros datos disponibles.

Sabemos también, que de los 45 empleados, 41 son camareros o cocineros y 4 se dedican a tareas de gestión de las cafeterías.

Una vez conocidos de que datos disponemos, o de cuales podemos disponer gracias a los recursos que tenemos, y si además existe personal disponible para encargarse de la gestión del análisis de datos, el siguiente paso es plantearnos qué nueva información sería atractiva obtener para nuestro negocio.

Ahora, para utilizar las herramientas y servicios que mejor se adapten a nuestro perfil, definiremos el objetivo a conseguir con la explotación de los datos recogidos.

La leche y el pan se compra a diario en estas cafeterías, por lo que en ocasiones, hemos percibido que existen días en los que sobran o faltan productos suponiendo un coste, o un coste de oportunidad respectivamente para la empresa. Por ello, hemos planteado analizar y conocer que productos se consumen cada día, y qué clientes lo han consumido. En esta ocasión, para analizar estos datos, se recogerán también información de la temperatura que hay en la terraza del establecimiento, condiciones meteorológicas, y festivos locales o nacionales entre otros.

El objetivo es minimizar los costes derivados de que sobren productos, o que de lo contrario, nos quedemos sin existencias y no podamos seguir atendiendo a los clientes. Para este objetivo, nos podemos plantear por ejemplo, el servicio que ofrece Nebinn, que puede analizar los datos obtenidos de distintas fuentes, como por ejemplo los datos recogidos con las tablets, información del TPV, o de la agencia estatal de meteorología para encontrar patrones ocultos en los datos.

Así por ejemplo, utilizando los servicios de Nebinn, podemos conocer qué tipo de productos se consumen cada día de la semana, usando algoritmos que predigan que cantidad de pan y de leche necesitaremos. Con esta información no sólo podemos ahorrar costes en la compra de género, sino que también podemos asignar el personal necesario pensando en las ventas de cada día. Añadimos que esto es extrapolable a cualquier producto ofrecido por la empresa, teniendo en cuenta el tiempo meteorológico o época del año junto con los productos consumidos, podemos

hacer una previsión de ventas anticipada para planificar las compras y personal necesario. Además, dispondríamos de datos como por ejemplo, cuáles son las elaboraciones más y menos consumidas, para así revisar nuestros precios y decidir si hacer más atractivos los productos menos solicitados, o quitarlos de la carta por la baja rentabilidad que nos ofrecen.

A lo anterior le podemos sumar solicitar el servicio de Commerce 360 o Mi comercio, ya que para los clientes del banco, el servicio es gratuito. De esta forma obtenemos información de los negocios ubicados en nuestra zona y sector, franjas horarias, códigos postales de nuestros clientes y de los clientes de la competencia, además de información demográfica. Así, si nos proponemos abrir una nueva cafetería en la misma ciudad, nos es muy útil conocer la procedencia de nuestros clientes y de los clientes de la competencia, para comprobar si la zona elegida tiene clientes potenciales, o clientes actuales de nuestras cafeterías. Además conociendo datos demográficos de la clientela nuestra y de la competencia, podemos aplicar estrategias para atraer al público menos asiduo a nuestras cafeterías.

Los datos ofrecidos en este ejemplo son reales, pero se mantendrá oculto el nombre de las cafeterías para proteger los datos de la empresa. Aún estos estudios no se han realizado en esta sociedad pero se han solicitado las respectivas demos de las herramientas, para comprobar sus beneficios y utilidad, para realizar así un estudio de viabilidad, y un Kick Off para comenzar a ponerlos en marcha.

6. CONCLUSIONES

En el presente estudio hemos hecho un recorrido por la historia del Big Data, hemos comprendido qué es este concepto del cual todo el mundo habla, y como se ha llegado hasta él. Hemos comprobado como las técnicas de Big Data son capaces de adaptarse a las pequeñas y medianas empresas y qué recursos hay disponibles para que puedan hacer uso de estas nuevas tecnologías.

Con todo lo anterior expuesto en el presente estudio, podemos concluir que si es posible el uso de técnicas de análisis Big Data para pymes.

Matizamos en que el esfuerzo que supone crear un nuevo departamento, invertir en ello, no es rentable para las pequeñas y medianas empresas. Hemos visto como existen recursos que no

necesitan de una gran inversión y que con un menor volumen de datos utilizan técnicas Big Data para procesar nuestros datos.

Si es posible que una pyme se beneficie del Big Data, pero no implantando una división o invirtiendo en nuevas tecnologías como pueda hacerlo una gran empresa, sino comprando el servicio a una compañía que se dedique a realizar este tipo de análisis.

Lo podemos comparar con cualquier otro servicio como pueda ocurrir por ejemplo, con el de contabilidad o marketing. Mientras una gran empresa tiene su propio departamento de marketing y personal cualificado que se encargue de lanzar una nueva campaña publicitaria, una pequeña o mediana empresa requerirá este trabajo a una agencia especializada para ello. No le es posible hacerse cargo de los gastos que supondría hacerlo ellos mismos.

Del mismo modo ocurre al usar servicios Big Data, una gran empresa como pueda ser por ejemplo un Banco, tiene su propio departamento de Big Data e invierte en nuevas tecnologías que le permitirán obtener nuevo conocimiento. Ahora bien, si pensamos en un pyme como pueda ser una cadena de 3 cafeterías, no puede asumir tales gastos ni tendría sentido hacerlo ya que nunca obtendría los ingresos para subsanarlos, por lo que puede beneficiarse contratando el servicio a otra empresa que ofrezca las soluciones que necesita, así sería posible obtener los beneficios que nos ofrece el Big Data sin asumir grandes gastos ni riesgos.

Las pymes pueden lucrarse con los distintos recursos que hemos ido viendo a lo largo del presente estudio, conocer su negocio y a sus clientes más de lo que son capaces de hacerlo actualmente. Esto otorga un gran valor, puesto que si a la cercanía con la que cuenta la pequeña empresa con el servicio ofrecido a sus consumidores, le añadimos obtener un mayor conocimiento sobre el negocio, permitirá abordar nuevos retos asumiendo menores riesgos y sin el temor de equivocación.

Como hemos demostrado a lo largo del estudio, estamos en la era digital, en la era de los datos, es imprescindible analizarlos e intentar conocer a nuestro consumidor y su comportamiento, para poder crear los servicios que mejor se adapten a sus necesidades. Hemos visto las distintas posibilidades de análisis que existen para las pymes y corroboramos que es necesario que comiencen a plantearse su uso.

6.1 CONCLUSIONES PERSONALES.

Este proyecto no ha sido un simple Trabajo de Fin de Grado, tampoco ha sido el cierre de una etapa ni el comienzo de otra, este estudio ha supuesto el nexo de unión durante un año, entre mis propósitos académicos y profesionales.

Me ha permitido conocerme, conocer mis inquietudes y poder afirmar, que realmente me encanta mi trabajo. Gracias a este estudio han fluido numerosas ideas, con las que me ilusionando pensando que en algún momento, y gracias a las técnicas de Big Data, se harán realidad.

Mi intención con la realización de este trabajo es llegar a todas las pequeñas y medianas empresas, mostrar un nuevo enfoque de negocio, hacerles conscientes del poder que tienen en sus manos y que son capaces de explotarlo. Quiero que gracias a este estudio se cuestionen, se hagan preguntas y apuesten por responderlas. Quiero que dejen de caminar solo por intuición, que conozcan a sus clientes y las necesidades que estos tienen, o comenzarán a tener.

La realización de este TFG me permitió “colarme” en el departamento de Big Data de una gran empresa, conociendo así su verdadero uso, los beneficios y dificultades que supone. Me ha permitido conocerlo e ilusionarme con nuevos proyectos. Esto es lo que quiero transmitir a la pyme española, que conozcan el gran potencial que tienen a su alcance, que se atrevan a soñar con nuevos proyectos, que sean conscientes del uso que les atribuyen las nuevas tecnologías y que apuesten por comenzar a hacerlos realidad.

7. BIBLIOGRAFÍA

Amazon Web Services (2017) “¿Qué es NoSQL?”, *Amazon*. Disponible on line: <https://aws.amazon.com/es/nosql/>

Artigas, C. (2016) “Big Data Week Madrid 216”, *Synergic Partners*. Disponible on line: <http://www.synergicpartners.com/la-evolucion-del-big-data-provocara-que-el-60-de-los-ninos-recien-nacidos-trabajen-en-puestos-que-hoy-en-dia-todavia-no-existen/>

Banafa, A. (2014). Small Data vs. Big Data : Back to the basics. *Analysis of Internet of Things (IoT) trends and technologies*. Disponible on line: <https://www.linkedin.com/pulse/20140703195144-246665791-small-data-vs-big-data-back-to-the-basics?trk=mp-reader-card>

Beller, M. J., & Barnett, A. (2009). Next generation business analytics. *Lightship Partners LLC*. Disponible on line: <https://www.slideshare.net/LightshipPartners/next-generation-business-analytics-presentation>

Cegos (2017). Big Data o Smart Data: El poder de la información. *Cegos University On line*. Disponible on line: <https://www.cegasonlineuniversity.com/big-data-o-smart-data-el-poder-de-la-informacion/>

Cox, M., & Ellsworth, D. (1997, October). Application-controlled demand paging for out-of-core visualization. In *Proceedings of the 8th conference on Visualization'97* (pp. 235-ff). IEEE Computer Society Press.

Dresner, H. (1993, May). Business intelligence: competing against time. In *Twelfth Annual Office Information Systems Conference, Gartner Group* (pp. 5-7).

Gartner, G. (2009). Gartner EXP Worldwide Survey of More than 1,500 CIOs Shows IT Spending to Be Flat in 2009. Disponible on line: <http://www.gartner.com/newsroom/id/855612>

GTAI (28 de noviembre de 2016). ¿Cómo deben las PYMES enfrentar la era del Big Data?, *Merca2.0*. Disponible on line: <https://www.merca20.com/deben-las-pymes-enfrentar-la-era-del-big-data/>

Hadoop Wiki (2016) “Apache Hadoop”. Disponible on line: <https://wiki.apache.org/hadoop/FrontPage>

Hernández, I. (18 de abril de 2017). “Big Data en Salud: entrevista a Nacho Hernández Medrano, PARTE 1”, *Instituto de Ingeniería del Conocimiento*. [Archivo de vídeo] Disponible on line: <https://www.youtube.com/watch?v=20tzdfBrNkk>

Hilbert, M., & López, P. (2011). The world’s technological capacity to store, communicate, and compute information. *science*, 332(6025), 60-65.

Index, C. V. N. (2017). Global Mobile Data Traffic Forecast Update, 2016–2021 Artículo Web. Disponible on line: http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.html?CAMPAIGN=Mobile+VNI+2017&COUNTRY_SITE=us&POSITION=Press+Release&REFERRING_SITE=PR&CREATIVE=PR+to+MVNI+white+paper

Laney, D. (2001). 3D data management: Controlling data volume, velocity and variety. *META Group Research Note*, 6, 70.

Lindstrom, M. (2016). Por qué el Small Data es el nuevo Big Data. *Knowledge@Wharton*. Disponible on line: <http://www.knowledgeatwharton.com.es/article/por-que-el-small-data-es-el-nuevo-big-data/>

Lith, A., & Mattsson, J. (2010). Investigating storage solutions for large data-A comparison of well performing and scalable data storage solutions for real time extraction and batch insertion of data.

López López J.C (2014) “La Moda del Big Data ¿En qué consiste en realidad?”, *El Economista*. Disponible on line: <http://www.eleconomista.es/tecnologia/noticias/5578707/02/14/La-moda-del-Big-Data-En-que-consiste-en-realidad.html>

Lyman, P., & Varian, H. (2004). How much information 2003?.

Martínez, J. (2016). Small Data: pequeñas pistas que aportan mucho valor a tu negocio. *Ainia centro tecnológico*. Disponible on line: <http://www.ainia.es/insights/small-data-pequenas-pistas-que-aportan-mucho-valor-a-tu-negocio/>

Pérez, P.F. (13 de marzo de 2017). Big Data: La clave de éxito de Juego de Tronos, Narcos o House of Cards. *TICbeat*. Disponible on line: <http://www.ticbeat.com/educacion/big-data-la-clave-del-exito-de-juego-de-tronos-narcos-o-house-of-cards/>

Spiegel, D. (4 de diciembre de 2014) “Un ‘software’ permitirá a la policía alemana predecir los crímenes antes de que ocurran”, *El Confidencial*. Disponible on line: http://www.elconfidencial.com/tecnologia/2014-12-04/un-software-permitira-a-la-policia-alemana-predecir-los-crmenes-antes-de-que-ocurran_551785/

Tanenbaum, A. S., & Van Steen, M. (2007). *Distributed systems: principles and paradigms*. Prentice-Hall.

Vodafone (2016). Análisis de la Digitalización de Autónomos y Pymes *PorqueTienesQueEstar.com*, 6. Disponible on line: <http://porquetienesqueestar.vodafone.es/wp-content/uploads/2016/09/informe-digitalizacion-pymes-autonomos-vodafone.pdf>

Zaharia, M., Chowdhury, M., Franklin, M. J., Shenker, S., & Stoica, I. (2010). Spark: Cluster Computing with Working Sets. *HotCloud*, 10(10-10), 95.