

RESEÑA

BIG DATA COMO ACTIVO DE NEGOCIO

Gemma Muñoz
y Eduardo Sánchez

Anaya Multimedia, Colección Social Media, 2019, 240 pp.



En el libro que a continuación se presenta, los autores, Gemma Muñoz y Eduardo Sánchez, introducen al lector en el universo del *Big Data*. Un concepto o nueva manera de entender el procedimiento en el que las empresas interactúan con sus clientes en el actual mercado dinámico.

La formación académica y trayectoria profesional de los autores, fundadores de la consultora digital El Arte de Medir, hace de esta obra una magnífica guía de cómo proceder en la recopilación, codificación, tratamiento y análisis de los datos para la mejora del posicionamiento de la empresa en el entorno

en el que actúan. Como los autores apuntan en la introducción de esta obra: «El *Big Data* nos permite obtener mucha información precisa sobre el cliente, sus costumbres y sus actos. Y nos ayuda a empujarlo a su próxima compra... Pero el dato sin contexto es solo un número». Ahí reside el carácter aplicativo de este libro, en una propuesta metodológica para trabajar con los datos de los clientes, la denominada por los autores metodología MAMBO.

El *Big Data* supone una innovación tecnológica de ayuda a las empresas para relacionarse comercialmente con su entorno, en especial con sus clientes, que ha venido para quedarse y que supondrá no solo la mejora comercial para algunas empresas, sino una cuestión de supervivencia de las compañías de determinados sectores. Este entorno en el que actúan las empresas es cada vez más dinámico y, por tanto, requiere un tratamiento ágil, casi al momento, de los datos disponibles para las compañías. Por esto, la buena utilización del *Big Data* se convierte en un aspecto determinante.

Este libro se estructura en nueve capítulos en los que se detalla una propuesta metodológica sobre cómo trabajar en el *Big Data*, con propuestas aplicativas en cada una de las fases propuestas por los autores. Así, en el primer capítulo se definen conceptos básicos de este entorno y se realiza una

explicación del origen de los tratamientos de datos, desde los inicios con *Inmon* y *Kimball* hasta la llegada del *Big Data*. Como se comenta en este capítulo, las empresas de cualquier sector siempre han dispuesto de datos, incluso más datos de los que eran capaces de procesar, lo que han cambiado ha sido el tratamiento ágil de estos, además de su centralización, almacenamiento y disponibilidad para las compañías. El segundo capítulo se centra en la importancia del capital humano para la correcta implantación de este sistema para que se convierta en un arma potente para la toma de decisiones. Para ello, es fundamental que toda la organización trabaje de manera conjunta y ordenada, siendo conscientes del objeto último y actuando cada uno en su ámbito.

Es en el tercer capítulo donde se aborda la metodología propuesta por los autores para desarrollar el *Big Data* en la empresa. Como paso previo se plantea la necesidad de saber el punto en el que se encuentra la empresa para su puesta en marcha, ¿está realmente preparada? Para ello se realiza un análisis de los principales modelos de madurez, como los modelos de Gartner, Cardinal Path, Brian Thopsey, Unilytics, Bersin & Associates o el modelo de madurez en analítica digital (El Arte de Medir, EAM). Finalmente, en este capítulo se comenta la propuesta de metodología para la implantación del

Big Data, la metodología MAMBO, denominada así por las siglas de cada una de las fases que la componen (Meditar sobre el negocio, Adquirir los datos, Manejar los datos, Buscar en los datos, Ordenar y visualizar). Los siguientes cinco capítulos del libro se dedicarán a la explicación y análisis de cada una de estas fases.

La primera fase de la metodología propuesta, a la que se dedica el capítulo cuarto, se basa en meditar sobre el negocio, es decir, en responder a una serie de cuestiones clave, como: ¿Qué recursos se disponen? ¿Qué tecnología se utilizará? ¿Qué datos serán necesarios? ¿Quién es el *stakeholders* del problema del negocio? ¿Cuáles disponemos? ¿Qué horizonte temporal se planteará? Según los autores, esta fase es fundamental, dado que empezar con mal pie, sin tener claros los objetivos, puede hacer que todo el trabajo posterior se quede en nada. El capítulo quinto se centra en la segunda fase de la metodología. En él se trata la tarea de adquirir los datos necesarios para el análisis, la identificación de las fuentes de datos de donde pueden ser extraídos y la evaluación sobre la idoneidad y suficiencia de los datos disponibles por la empresa, ¿son suficientes y de calidad? En el capítulo también se indica la necesidad de realizar una limpieza de los datos adquiridos, comentando herramientas eficaces para tal efecto, como el *OpenRefine*.

Es en la tercera fase de la metodología, capítulo sexto, donde se aborda el manejo de los datos, la fase más amplia del proceso. En este capítulo se determinan los métodos y técnicas para el desarrollo de la tarea, diferenciando entre el análisis en tiempo real del proceso de datos por bloques y el estudio de datos cualitativos del análisis de datos cuantitativos. Así se tratan técnicas como la regresión lineal, el análisis de correlaciones o el análisis semántico. Además, se trata la importante tarea de la preparación previa de los datos (limpieza de datos, tratamiento de datos perdidos o mal grabados y eliminación de valores extremos que pudiesen distorsionar los resultados).

El capítulo séptimo se centra en cómo extraer respuestas a las cuestiones planteadas de los datos trabajados. La clave de este proceso se encuentra en la búsqueda de *insights* y en el resumen de los resultados extraídos para que puedan ser de utilidad en la toma de decisiones. El objetivo de esta fase es predecir acontecimientos y proponer acciones de mejora en la empresa. Así, es importante diferenciar tres tipos de análisis en esta fase: *i*) análisis de diagnóstico, identificar las causas de los acontecimientos acaecidos; *ii*) análisis predictivo, establecer qué pasará en el futuro si se sigue actuando de la misma manera; y *iii*) análisis prescriptivo, qué se debe realizar para cambiar

las cosas. Este último análisis será el más complicado y necesitará un conocimiento profundo de la empresa y del sector y del entorno en el que actúa la compañía. Aunque con frecuencia el análisis de datos se queda en la fase anterior, los autores subrayan la importancia de una quinta fase a la que dedican el capítulo octavo. En este capítulo se trata cómo ordenar y visualizar los datos, siendo esta una fase determinante para terminar de diseñar los *insights*, mediante la colaboración con los *stakeholders*. El capítulo se completa con la identificación de las principales herramientas de visualización.

El último capítulo del libro se dedica al análisis de fuentes de datos adicionales que están a disposición de la empresa, como los datos ocultos (*dark data*), disponibles de diversas formas y no siempre tenidos en cuenta. Además, se comentan otros métodos de extracción de datos aún no explotados o definidos en su totalidad, como el *blockchain* o *streaming* de datos o disposición de datos en tiempo real. Como epílogo del libro, los autores realizan una serie de conclusiones a modo de reflexiones finales sobre el estado actual del *Big Data* y de su futura y posible evolución, así como el cambio que esta metodología supone para el ámbito empresarial.

Rodrigo Medina Viruel
Universidad de Jaén