

## La ciencia de los datos y el conocimiento en la empresa

Fernando Piera <sup>1</sup>

### Resumen

Nos encontramos en el tiempo de la explosión de los datos. Hasta se está hablado de la ciencia de los datos. Pero los datos no son más que el “átomo” de la información y esta a su vez es el componente del conocimiento como es bien conocido y se trató en uno de los primeros cuadernos de esta colección. La información aparece cuando los datos son manipulados con un objetivo determinado, lográndose entonces organizar con los datos una información. La combinación de informaciones sobre un objetivo determinado nos va a proporcionar el conocimiento que se requiere para la toma de decisiones en las empresas.

Puede que no fuera el inventor de la ciencia de los datos, pero el profesor Hal Varian de la Universidad de Berkeley en California (EEUU) pero si quien le dio credibilidad cuando en Octubre de 2008 manifestó que la profesión del futuro seria la que tuviera la habilidad de hacer posible que los datos fueran comprendidos, de procesarlos, de extraerles valor, de visualizarlos, de comunicarlos, todo lo cual será una tarea tremendamente importante en las décadas siguientes. Y no solamente para el profesional sino también para la educación en sus distintos niveles, desde la enseñanza primaria a la enseñanza universitaria. Y esto porque ahora se dispone de datos libres y ubicuos. Añadió entonces el profesor Varian que la consecuencia era entender el significado de los datos y extraer su valor.

El profesor Varian siguió explicando que la ciencia de los datos era como una disciplina estadística, pero señalando que hay más que el puro numero aislado, que la tarea esencial es de visualización, comunicación y utilización de los datos. La estadística es parte de ello, pero solo una parte. También se desea ser capaz de visualizar los datos, comunicarlos, utilizarlos de manera efectiva.

El profesor Varian también opinaba que estas capacidades de poder acceder, entender y comunicar el conocimiento que se puede obtener del análisis de los datos, seria de extrema

---

<sup>1</sup> Socio senior de ATI

importancia para que los gestores y directivos de las organizaciones, de las empresas e instituciones, puedan acceder y comprender los datos directamente ellos mismos.

Recientemente se está escribiendo mucho para enmarcar los límites de la ciencia de los datos, la mayoría por estadísticos, los cuales han migrado hacia el artículo sobre analítica de la revolución de los negocios de David Champagne publicado en Enero de 2011, derivado a su vez de lo publicado por el profesor Drew Conway, científico de política y datos de la Universidad de Nueva York, para conceptualizar la ciencia de los datos como un lugar de confluencia de la ciencia estadística, Hacking informático y experiencia substancial en gestión de organizaciones.

La capacidad de improvisar soluciones de integración de datos utilizando sistemas operativos, bases de datos y lenguajes de programación distingue claramente la ciencia de los datos de la tradicional investigación estadística. Siguiendo al bioestadístico Michael Elashoff, el término “científico de datos” es realmente más de un reconocimiento que la gente en este campo necesita múltiples tipos de conocimientos. Reconoce el hecho de que mirando a los datos requieren más que habilidades analíticas.

El trabajo de interpretar datos para ayudar en la toma de decisiones, en las empresas e instituciones, parece arrancar hace 5000 años con los burócratas y comerciantes sumerios en el Creciente Fértil del Medio Oriente. Pero tratar con el tamaño enorme y complejidad de los conjuntos de datos modernos requiere un nuevo conjunto multifaceta de habilidades computacionales, estadísticas, comunicativas, denominada “ciencia de los datos”. Los científicos de los datos están entre los profesionales más buscados hoy en día y la necesidad que se tiene de ellos crecerá según el trabajo crezca en complejidad e interconexiones en las organizaciones.

Cuando en TI se habla de datos, se está hablando sobre los elementos físicos almacenados en sistemas. Cuando marketing habla de datos se está refiriendo a los totales y cálculos resultantes de los análisis.

Las reglas de los datos son algoritmos. Cuando se aplican las reglas a datos, estos están más alineados a la segmentación y estatus que serían input para la interacción personalizada con clientes. Las reglas de los datos no se refieren a los comerciantes de información.

## **Pero ¿Qué aspecto tienen y donde se pueden encontrar estos científicos de los datos?**

---

La función del científico de los datos es, en breve, hacer que los datos tengan sentido, facilitar una comprensión clara del universo de la organización a través del análisis de los datos, ayudando la mejora del proceso de toma de decisiones y apoyando al liderazgo.

Los mejores científicos de los datos también deben interesarse por el descubrimiento de nuevas visiones de los datos. Serán creativos en su aproximación a la identificación y resolución de problemas. Su experiencia profesional se encuentra en los tres pilares:

- 1) Profundo conocimiento teórico de estadística y conmutabilidad;
- 2) Conocimiento práctico de variadas herramientas de ciencia de los datos y la capacidad de comunicar efectivamente cuando es necesario;
- 3) Una capacidad para comunicar de manera efectiva con la gente que no tenga conocimientos técnicos sobre temas técnicos muy complejos.

Pueden utilizar un surtido variado de las herramientas existentes de análisis de datos y también tener el conocimiento estadístico y habilidades de programación necesarias para desarrollar sus propias herramientas. Muchos problemas de análisis pueden ser resueltos utilizando las herramientas existentes, pero el científico de los datos también debe estar en condiciones de desarrollar las herramientas que necesite para atacar nuevos problemas que se le puedan plantear.

Los científicos de los datos deberán conocer las muchas variedades de datos útiles y conocerán como recoger datos importantes del tipo correcto para resolver un problema determinado y como estructurar y limpiar los datos, de manera que puedan ser analizados de manera adecuada. Los datos, en el mundo real, nunca han sido facilitados en una bandeja de plata y el científico de los datos debe saber como localizar los datos correctos y prepararlos para su utilización.

En materia de comunicación, y además de lo mencionado anteriormente, deberán estar en posesión de sólidas capacidades de escuchar, hablar en público y habilidades para presentaciones visuales.

Los métodos de comunicación y análisis y sus resultados pueden parecer problemáticos ya que no hay una única respuesta a cada pregunta que se formule, pero también existen importantes dificultades y limitaciones basadas en complejas asunciones estadísticas y la naturaleza de los datos utilizados.

Los científicos de los datos deben estar preparados para ayudar a los que toman las decisiones en las organizaciones, sean empresas o instituciones, en la tarea de refinar y

clarificar sus ideas sobre los problemas que se les plantean a diario. Igualmente deben poder ayudarles a convertir grandes y globales problemas de los negocios en cuestiones específicas y significativas a los que los datos pueden facilitar una visión clara.

El científico de los datos, al cual también se le podría denominar experto en la manipulación y gestión de datos, estará cómodo trabajando con individuos de diversa formación y con variadas habilidades. Un proyecto de ciencia de los datos reúne a directivos con estadísticos, ingenieros de software, analistas de estrategia, expertos del asunto tratado y otros.

Los científicos de los datos deben ser capaces de trabajar de manera efectiva con todas las personas interesadas y coordinar su trabajo para alcanzar soluciones prácticas y eficientes. Ellos pueden acelerar la utilización de nuevas aplicaciones y determinar el modo y manera de acercarse adecuadamente a la colecta, análisis e interpretación de datos-

Aunque trabajen para una única organización se encontraran con un surtido variado de problemas que requieren su ayuda. Ellos tienen que ser capaces de obtener una clara comprensión de la naturaleza única de cada problema nuevo y no limitarse a fiarse de un concepto abstracto de cómo deberían ser los datos.

### **Los desafíos de la Ciencia de los Datos**

---

La ciencia de los datos trae un conjunto de desafíos específicos a la comunicación en los negocios. Como se puede explicar la evidencia estadística y los resultados analíticos sin simplificar en exceso o crear confusión. Los estudiantes necesitan aprender a elaborar los resultados de forma coherente, como explicar las asunciones estadísticas y los vacíos de manera clara y como las visualizaciones que facilitan una perspectiva profunda y no son solo gráficos bonitos. La única manera de aprender estos conocimientos es practicándolas al mismo tiempo que se aprenden las técnicas inherentes. Finalmente, un componente crítico de cualquier programa educativo de calidad de ciencia de los datos es una especie de componente de experiencia práctica, tanto si se trata de un proyecto estudiantil, de un programa interno o de una guía práctica.

Hasta que se trabaje con problemas de un mundo real de ciencia de los datos los estudiantes no comprenderán totalmente la manera de formar actividades centrales de ciencia de los datos que no puede ser enseñada en un aula académica: Luchando por definir correctamente el problema a analizar, lidiando con las complejidades e inconsistencias reales y satisfaciendo las modas a los clientes no técnicos ni académicos.

Dentro de estos desafíos se encuentra el referente a los datos no estructurados que plantean serias dificultades para su gestión y su manejo y tratamiento, la eliminación de datos repetitivos o duplicados y el logro de la debida calidad en los datos que se gestionan.

## **Una nueva ciencia y una nueva profesión**

---

Romper y abrir el espacio entre todos los silos y modelos de datos desconectados y crear para describir datos. Esta es la idea de Suzi Stover por la que crea un nuevo marco para describir datos que derriban los muros, separaran los datos de los modelos y almacenes que los confinan, convirtiendo todo ello en una abstracción, una manera de profundizar en todos los datos, de manera semántica y con las interfases disponibles sin disturbar o perder nada de yos en el proceso. Desde otro punto de vista el encaje es simplemente tan grande que existe principalmente para perpetuarse a si mismo. .

El punto es que el científico de los datos no es meramente un nuevo nombre para la gente con grandes habilidades analíticas como a algunos pueda parecer, sino mas bien una designación para la gente que, además de capacidad analítica, tiene un amplio conocimiento de las diversas áreas importantes para lidiar con los problemas complejos de la vida real de la ciencia de los datos. Tratar con el tamaño, la velocidad, la diversidad y la complejidad de los modernos conjuntos de datos requiere mucho más que ser experto en estadística y herramientas de software comercial. Un científico de los datos debe saber sobre representación de datos, ingeniería del software, diseño de visualización de datos, y procesos de los negocios, al igual que ser capaz de contar la “historia de los datos” basándose en rigurosos análisis cuantitativos.

Este científico de los datos no es un ser de fabula, sino un profesional serio que comprende la globalidad del ciclo de vida de un problema de ciencia de los datos y también tiene conocimientos y habilidades en diversos y específicos aspectos de la ciencia de los datos. Es decir, hay muchos tipos de científico de los datos, unos pueden ser mejores en estructurar y limpiar datos formalizados, poniéndolos en una estructura uniforme que puede ser analizada de manera efectiva, y otros pueden ser expertos en la elaboración de modelos estadísticos y otros pueden ser unos sólidos desarrolladores de sistemas de software de alto rendimiento para análisis de datos. Pero todos ellos comprenderán la necesidad de normalizar los datos, la asunción hecha por diferentes métodos estadísticos y inconvenientes inherentes al desarrollo de sistemas informáticos flexibles igual de cómo aprender rápidamente un oficio o un negocio, formularlo como una tarea de ciencia de los datos, y comunicar los resultados de manera efectiva a los que toman las decisiones en la organización.

Estos científicos de los datos pueden carecer de la productividad de superhombres o del impacto de una estrella del Bel Canto. Si resulta posible encontrar uno de estos expertos que resultan indispensables para el desarrollo de esfuerzos importantes en materia de ciencia de los datos, incluso los que están en equipos multidisciplinarios.

Como cualquiera que haya probado a constituir un equipo a partir de estos variados profesionales como los estadísticos, ingenieros de software, analistas de negocios, o especialistas de comunicación, sabe, resulta increíblemente difícil establecer puentes que resuelvan los gap culturales y lingüísticos entre ellos y desarrollar un grupo coherente que pueda trabajar conjuntamente de manera efectiva.

El científico de los datos sabe como pensar y como hablar como un estadístico, como un ingeniero de software, como un analista de negocios, como un especialista de comunicaciones y así pudiendo puentear estos gaps y asegurar un suave proceso de ciencia de los datos. De hecho tal equipo es como trabajar de la mejor manera cuando la mayoría de sus miembros son científicos de los datos. Solo con unos pocos y sólidos especialistas involucrados es suficiente. Teniendo en cuenta que cada experto, cada científico de los datos suele tener enfoques diversos en su aproximación a los problemas, resulta evidente que un equipo de varios expertos puedan abordar los problemas complejos de la ciencia de los datos.

No se debe creer la fantasía de que los científicos de los datos son figuras míticas y se debe, definitivamente no ser negligente en el desarrollo de las capacidades de la ciencia de los datos. Resulta difícil imaginar una organización, que hoy en días, no se beneficie de un serio programa de ciencia de los datos. Y la mejor y más fácil manera de realizar un tal esfuerzo de encontrar científicos de los datos es crearlos. No unicornios, no superhéroes, pero simplemente científicos de los datos. . No hay mitos, `pasean por la tierra hoy en día su numero esta creciendo y todos los necesitamos.

### **La ciencia de los datos y el big data**

---

Los gurús del big data han dicho que la calidad de los datos no es importante para big data. Bueno es suficiente y basta. Sin embargo los titulares de los negocios se quejan aun sobre la falta de calidad de los datos. De hecho, en una encuesta de Forrester sobre los profesionales de inteligencia del negocio en los clientes la capacidad de integrar datos y gestionar calidad son los dos factores para mantener un cliente.

Big Data esta transformando la manera en que la ciencia e incluso los negocios, se llevan a cabo. La revolución del big data esta hacienda que los limites sean una cosa del pasado. Piénsese que este nuevo paradigma “análisis de totales” mejor que simples habilidades de triturar grandes conjuntos de datos-. A lo largo del tiempo, según evoluciona el mundo hacia aproximaciones masivamente paralelas con herramientas tales como Hadoop, se `podrá realizar análisis de 360 grados auténticos. Por ejemplo, según una población mayor va entrando en las redes sociales y lleva su vida en forma publica en línea, todos tendremos una inteligencia de mercado detallada y comprensiva de cada elemento demográfico disponible como si fuera un público recurso-

Según el precio del almacenamiento y de la banda ancha continúe su bajada los científicos de los datos podrán mantener conjuntos completos de poli estructurados informaciones bajo sus algoritmos microscopios mejor que tener que fiarse manualmente de muestras, subconjuntos u otras variantes.

La idea original era ensamblar un conjunto de datos a partir de informaciones disponibles de manera inmediata en la Web que fuera apropiada para una visualización interactiva complementada por una modelización de análisis predictivo.

La ciencia de los datos tiene históricamente que mal contentarse con meras muestras de datos. Pocos científicos de los datos han tenido el lujo de ser capaces de acumular petabytes de datos en alguna variable relevante en cada entidad de la población objeto del estudio.

### **Situación de la formación de los profesionales de la ciencia de los datos**

---

Nos encontramos en 2014 con una gran demanda de profesionales que este en condiciones de lidiar con los problemas que plantean los volúmenes ingentes de datos y lograr que proporción un conocimiento indispensable para la toma de decisiones en las organizaciones, en las empresas.

Ya se ha definido anteriormente las la función que estos expertos o si queremos “científicos de los datos”, o ¿expertos de los datos”, incluso “gestores de la información” (no CIO que tiene otras funciones) y para poder llevar a cabo estas funciones se requiere un proceso formativo que esta comenzando a desarrollarse en los EEUU (su lugar de origen), pero también las universidades europeas están entrando en este tema, incluso alguna universidad privada española intenta abrase camino en esta naciente disciplina.

En síntesis, las universidades estadounidenses más avanzadas están poniendo en el mercado académico, masters de 4 semestres, alcanzando los 36 créditos y con inicio en 2014 y con un contenido que cubre materias como:

- análisis de datos, modelización estadística, ingeniera de datos, bases de datos relacionales, SQL y non SQL, manipulación de datos a es escala de big data, algoritmos para minería de datos, calidad de los datos, remedios y operaciones
- análisis predictivo
- Servicios para empresas gestionadas en base a datos
- Rendimiento tradicional de la gestión que es lo que ha mantenido la BI 20 años en boga.
- Exposición de datos y analítica como producto. Las redes sociales ofrecen unas muy atractivas ilustraciones de este modelo.
- Los conceptos fundamentales de ciencia de los datos.

- Una sólida y genérica guía para profesionales del análisis,
- Diseño de la metodología de generación/recolección de datos.
- Métodos de exacta y aproximativa inferencia en modelos gráficos, así como las herramientas útiles para la gestión de los negocios.

Al mismo tiempo parece necesario el aprendizaje de técnicas para modelos gráficos y predicciones estructuradas, así como una introducción a la visualización de datos, inferencia y representación

Los modelos gráficos, la inferencia causal, y otros tópicos avanzados en aprendizaje de estadística por máquinas hay que tenerlos también en cuenta. Aprender experimentalmente con herramientas de visualización de datos, incluidas herramientas de código abierto resulta otra de las áreas que no se puede olvidar.

En relación con el tema formativo es central tanto interna como externa la validación de las requisitorias de ciencia de los datos. Por otra parte el tema práctico y de proyectos es muy importante. Un experimento aleatorio de series temporales con un grupo de control natural diseñado resulta mucho más contextual y promueve una mayor confianza que un análisis correlacional único. Y los científicos de los datos deben ser conscientes de ser demasiado crédulos con sus datos, tomando precauciones de no sobresaturar los modelos.

Facilitar experiencias operativas y prácticas de tipo real en los más importantes algoritmos de regresión lineal, regresión logística son materias que bien en forma básica u optativa tienen que aparecer en un plan de estudios de suficiente calidad.

Se recomienda aprender como pensar sobre datos y como comenzar a extraer significado de los mismos. Resulta imprescindible cubrir el tema de big data abarcando metodologías y herramientas para extraer conocimiento de enormes conjuntos de datos. Estos métodos incluyen aprender hashing, embebido de clases, bases de datos distribuidas, esquemas para lectura de mapas, programación CUDA GPL programación y aplicaciones.

Sería conveniente añadir una experiencia práctica con herramientas de big data como mapa de reducción, programación paralela, aprendizaje en línea y métodos de hashing.

Los métodos de análisis automáticos, que los científicos de los datos puedan utilizar para crear nuevo conocimiento a partir de flujos enormes cantidades de datos no estructurados de diversas procedencias son otra de las materias a tener en cuenta.

Además hay que tener en cuenta el análisis automático de conjuntos masivos de datos incluyendo nuevas ramas de genomics, bioquímica, aeronomía y cosmología, física de altas energías, y neurociencia, especialmente imágenes cerebrales. Algunas áreas de las ciencias



sociales y de las humanidades están ya siendo influidas por análisis automático de datos, incluyendo derecho, economía, sociología, ciencia política e historia.

Y no se puede olvidar el estudio de la calidad de los datos de la información obtenida de los mismos, así como el tratamiento de la utilización de los datos no estructurados y su conversión en estructurados.

La prensa norteamericana ya ha abordado el problema de la escasez de especialistas en la ciencia de los datos. El diario US Today ha dicho que solo hay el 20% de científicos de los datos necesarios y demandados.

Herramientas de diseño, manipulación y gestión de datos.

Solo por mencionar alguna de las universidades estadounidenses que están abordando el problema se puede mencionar la Universidad de Nueva York y la Universidad de California en Irvine, California.

Con la colaboración de:





*Confederación Española de  
Directivos y Ejecutivos*

World Trade Center

Moll de Barcelona, s/n Edificio Este 1ª planta  
08039 Barcelona (Spain)

Tel. +34 93 508 83 20  
Fax. +34 93 508 83 21

Correo: info@directivoscede.com

Acceda a nuestra web:

[www.directivoscede.com](http://www.directivoscede.com)

*La Confederación Española de Directivos y Ejecutivos (CEDE), fundada por la Asociación Española de Directivos (AED) y la Asociación Española de Ejecutivos de Finanzas (AEEF), tiene como objetivo principal agrupar a distintas asociaciones de directivos al objeto de dotarlas de una representación en consonancia con la relevante función que desempeñan en el ámbito socioeconómico, así como estimular sus actividades y coordinar acciones en beneficio de las entidades confederadas y de todos sus socios individuales. Desde su fundación en 1997, se han integrado en la Confederación otras asociaciones de directivos pertenecientes a diversos sectores y ámbitos territoriales. En la actualidad CEDE agrupa 46 entidades y más de 120.000 directivos, por lo que está alcanzando una representatividad significativa, que continuará incrementándose en el futuro con la incorporación de nuevas entidades.*

## Sobre la Comisión de Gestión del conocimiento empresarial

Si aceptamos que hoy en día el conocimiento es un factor esencial para el desarrollo empresarial y que los activos intangibles son, cada vez más, la parte creciente del valor y de la capacidad de generar resultados sostenibles, se considera de especial relevancia incorporar en el programa general de actividades de CEDE la actuación de un Observatorio sobre Gestión del Conocimiento Empresarial, Capital Intelectual e Innovación, que será gestionado por esta comisión de trabajo.

La comisión de trabajo considera prioritario enlazar cualquier iniciativa que encaje con la mejora de la gestión empresarial y, específicamente, con el incremento de la calidad de las funciones directivas y ejecutivas. Por tanto, el enfoque de sus actividades será eminentemente práctico, obviando perspectivas excesivamente académicas.

También sus actuaciones se enmarcan en un contexto global, alineadas con acciones y trabajos realizados en el marco de la Unión Europea.

En cuanto a los objetivos de trabajo, principalmente son:

a) Seguimiento de las principales corrientes de opinión y avances teórico-prácticos en materia de activos intangibles, capital intelectual, gestión de la información, inteligencia empresarial y tecnologías de gestión del conocimiento.

b) Acuerdos con fundaciones y entidades de estudio nacionales e internacionales, cuyo ámbito de investigación merezca ser seguido, conocido y divulgado por CEDE.

c) Difusión en los medios virtuales de CEDE de las mejores prácticas y los avances más significativos en Gestión del Conocimiento, Capital Intelectual e Innovación.

d) Creación de plataformas de debate y aprendizaje, de modo que pueda mantenerse una Red de intercambio de Conocimiento entre las asociaciones de CEDE.

Puede conocer más sobre nuestra actividades y cuadernos en la web de CEDE o solicitando información en el email [info@directivoscede.com](mailto:info@directivoscede.com)