

Productores de datos

Te despiertas por la mañana. Coges el móvil, miras la hora. Las nueve, hora de levantarse y desayunar. Miras rápidamente WhatsApp/telegram/e-mail y respondes a los mensajes que tienes pendientes. Felicidades. En ese preciso momento los «señores de traje» de WhatsApp/telegram/Google conocen a qué hora te despiertas y por supuesto pueden conocer la ubicación de tu casa con un grado de exactitud bastante alarmante.



A lo largo de un día le enviamos numerosos mensajes a través de diferentes aplicaciones móviles. Ahora bien, ¿a quién pertenecen estos datos? Por una parte podrías pensar que todos estos datos deberían ser tuyos, ya que tú los has producido. Pobre ingenuo. Como dicen en el juego Zero Wing: «**All your base are belong to us**». No nos equivoquemos. Estos no son los únicos datos que producimos cada día. Cada día nuestra actividad en facebook/twitter/google + o cualquier otra web es trackeado, almacenado y analizado para la posterior obtención de conclusiones.

Seguramente, ninguno de los que estáis leyendo este artículo os habréis dado cuenta pero de este y cualquier otro artículo escrito en internet se pueden sacar diferentes conclusiones muy diferentes a las propuestas por el propio artículo. Por ejemplo, en este escrito hasta este párrafo se han escrito 153 palabras, 972 caracteres y la palabra más usada (descartando preposiciones, disyuncion

es...) es: «**datos**». Puede parecer una tontería, pero de esta forma se podría calcular el promedio de palabras escritas por una persona en un post, de forma que si un día escribo un post excesivamente largo o corto se puede deber a diversos factores como el interés que pueda tener sobre el determinado tema o mi estado de ánimo. Por otra parte, se podría analizar un artículo para ver si dicho artículo ha sido «influenciado» por otros artículos y de esta forma ver si el autor lo ha escrito «porque tenía que escribir algo» o si es algo que ha salido de él.

Finalmente concluir, que la próxima vez que vuestra madre, novia o amigo listo de turno os diga «os pasáis todo el día sin hacer nada, ahí tumbado, móvil en mano», podréis responderle «estoy produciendo... datos». También recalcar la importancia que tiene a quién pertenecen estos datos y las

diferentes conclusiones que se podrán obtener a partir de ellos.

Realmente... ¿Necesitamos los sistemas de información?

Hoy en día vivimos rodeados de dispositivos electrónicos los cuales recogen datos de nuestro comportamiento, las acciones que realizamos... Todo ello para que las empresas puedan tener una información más exacta sobre sus clientes y de esta forma prestarles un mejor servicio. Junto con esto podríamos intentar preveer qué querrá un cliente antes de que le surja esta necesidad. De hecho, Amazon lo está intentado. Para ello quieren saber cuándo tendremos la necesidad de comprar un producto determinado para así enviarnoslo a nuestras casas. Imagínatelo, te apetece comer tomate y según se te ocurre, tienes en la puerta de tu casa a un señor con una caja de tomates. Ahora bien, esto no ha sido siempre así. Si miramos hacia atrás, exactamente hace un siglo, veríamos que todo esto parecería algo impensable. ¿Por qué?. Simplemente es que no había esa necesidad. Los sistemas de información son una necesidad actual debido a los cambios que han ocurrido gracias a la evolución de las diferentes tecnologías, los avances en logística... Por ejemplo, antes era impensable transportar en un día un producto desde un América hasta Europa.

A continuación se exponen algunos ejemplos de por qué hace 100 años no necesitábamos un sistema de información para que nuestro negocio funcionara.

Hace 100 años las empresas no eran del mismo tamaño del que son ahora. Frente al modelo actual en la que una gran empresa puede llegar a dar servicio a más de un país, antiguamente estas daban servicio como máximo a un sólo país. La gran mayoría de «grandes empresas», eran una tienda situada en la esquina de una calle donde el hijo de un agricultor local vendía sus productos. Tras esto, nos quedan otra incógnita, si esa empresa era «grande de verdad», ¿cómo llevaban a cabo la contabilidad de un empresa? Este problema se explica en el siguiente párrafo.

Otro punto a comparar es la cantidad de información a procesar. Pongamos un ejemplo sencillo: las ventas del día de una mediana empresa. Antes, para llevar las cuentas de una empresa bastaba con tener papel y lápiz. Actualmente, las ventas que hace una pequeña empresa no tiene por qué ser sólo lo que se vende en una pequeña tienda. Muchas empresas venden a través de internet y estos datos deben estar centralizados junto con las ventas de una pequeña tienda.

Lo explicado anteriormente es un pequeño ejemplo de por qué actualmente es impensable un mundo sin sistemas de información. Imagínate por un momento que una empresa como Telefónica no sabe quiénes son sus clientes. ¿Estaría bien, e? De esta forma, muchos de sus «clientes» no tendrían que pagarles a fin de mes, pero por el otro lado, esos «clientes» recibirían muchas más llamadas de

otras compañías para ver si te «olvidas» de Telefónica (o la compañía que tengas contratada) y decides ser su nuevo cliente. Seguro que esta última opción nos gusta menos a todos.

Finalmente me gustaría lanzar una pregunta al aire para la reflexión de los lectores. ¿Os imagináis un mundo sin sistemas de información?.

Empresas que ya combatieron el fraude

En este post se cuentan las experiencias de algunas empresas para combatir el fraude.

PayPal

Esta empresa necesita evitar los casos de fraude a través de una aplicación que pueda hacerlo en tiempo real. Este caso de Big data se debe a que ellos manejan información a través de PC, teléfonos, como medio de pago estándar en miles de sitios webs y muchas plataformas más que causan que l



os casos de fraude se acumulen cada vez más. Por este motivo es que necesitaban alguna herramienta para poder predecir dónde se podría llegar a cometer fraude antes que este comience a realizarse.

Con el objetivo de lograr paliar el fraude, realizan un análisis continuo y en tiempo real en grandes conjuntos de datos y aplican patentes de Google con la indexación de la web para transferirlo al problema de calcular el riesgo de fraude para los consumidores o comerciantes individuales

CSM-ACE

En la CSM-ACE de 2012, Neil Meikle (PwC) presentó el siguiente caso de estudio.

Una productora de televisión y compañía de emisión descubrió un fraude de facturación. El cliente sospechó de otras instancias de facturación falsa durante dos años. Durante este tiempo se realizaron 200.000 transacciones y 9.500 ventas. Estas transacciones exhibían un gran rango de valores y se desconocían qué transacciones eran fraudulentas. La solución a este problema se logró haciendo clustering de los suministradores.

Para ello se agrupó a los proveedores basándose en sus características. De esta forma se pudo localizar los diferentes valores “anómalos”. Tras este análisis se encontraron 42 vendedores anómalos de los cuales dos fueron confirmados como fraudulentos

Usando el «Big Data» para analizar riesgos

En un mundo que aumenta de complejidad y aumenta la demanda, la capacidad para capturar, acceder y utilizar el “Big Data” determinará el éxito en el análisis de riesgos. Este análisis masivo de datos permite crear modelos de riesgo de forma que se aumentará y mejorará las respuestas que se entregarán dando lugar a prever anteriormente los riesgos y ahorrar dinero. En palabras de Jason Hill “Time is critical in the new world of risk management. If you can react to a risk faster, you have a competitive advantage”.

La época de los warehouses ha terminado y se empieza a hablar de “lagos de datos” en los cuales se guarda la información en su formato original. De esta forma, un banco puede saber el riesgo que tiene dar una hipoteca a una determinada persona basándose para ello en su gasto, hábitos de compra, historial previo de préstamos, etc. Además, se puede identificar lo antes posible actividad sospechosa antes de que actos devastadores creen daño en una empresa.



Imagen 1: Descripción de las oportunidades que ofrece las tecnologías “Big Data” para el análisis de riesgos

El análisis de riesgo se enfrenta a nuevos desafíos. En respuesta a la actual crisis, se están obteniendo cada vez información más detallada y por ello informes más sofisticados. Actualmente hemos visto muchos casos de blanqueo de dinero, a través del “big data” se puede monitorizar y modelar estas transacciones de forma que se pueden detectar en tiempo real. En la imagen de la izquierda se muestra algunas de las oportunidades que ofrece el «big data».

Ejemplos de uso del big data

Credit Risk

Tras el inicio de la crisis, las instituciones financieras deben tener un mayor conocimiento sobre los diferentes clientes. Por ejemplo, a través del análisis de datos pueden desarrollar indicadores más robustos los cuales les asesoren en el dominio del crédito. Las nuevas fuentes de datos se pueden usar para obtener mayor información del comportamiento de los usuarios. De esta forma, se pueden apreciar diferentes alarmas en los usuarios simplemente visualizando su comportamiento y sus acciones a través del tiempo.

Anti-blanqueo de dinero

Viendo diferentes acontecimientos en la economía/política actual somos capaces de observar que las "altas esferas" blanquean gran cantidad de dinero y las formas tradicionales de detectarlo se basan en un anticuado sistema de reglas y en análisis descriptivo para procesar la información estructurada. Es por ello, que tiene sus propias limitaciones.

Los análisis de big data pueden mejorar los procesos de detección de blanqueamiento de dinero. Este nuevo enfoque permite el análisis estadístico de datos estructurados, visualización de ellos y minería de texto de estos datos. Estos enfoques proveen un vistazo rápido a los diferentes vínculos ocultos existentes entre transacciones y cuentas y descubrir patrones de transacciones sospechosos a través de técnicas de Machine Learning.

Riesgo de mercado

La cuantificación de entregar un crédito es un riesgo el cual es cada vez más complejo. Para ello es necesario tener en cuenta diferentes componentes los cuales requieren analizar una gran cantidad de datos. Para ello, los grandes bancos suelen tener escenarios como los de "Monte Carlo". Este análisis debe ser muy rápido y para ello los ordenadores suelen procesar esta información en memoria, ya que en este tipo de riesgo, el volumen de información y la velocidad son factores clave.

[Introduciéndonos al Big Data](#)

Big data is like teenage sex: everyone talks about it, nobody really knows how to do it, everyone thinks everyone else is doing it, so everyone claims they are doing it..

Con esta frase empezó Dan Ariely una de sus conferencias. Lo que quiere decir es que hay numerosas empresas que dicen que ellos realizan tareas de análisis de datos para obtener mejores resultados que la competencia en determinados campos, pero, ¿realmente están haciendo “big data”? En la Imagen 1 se muestra la opinión de los directivos sobre qué es esta nueva “buzzword”.

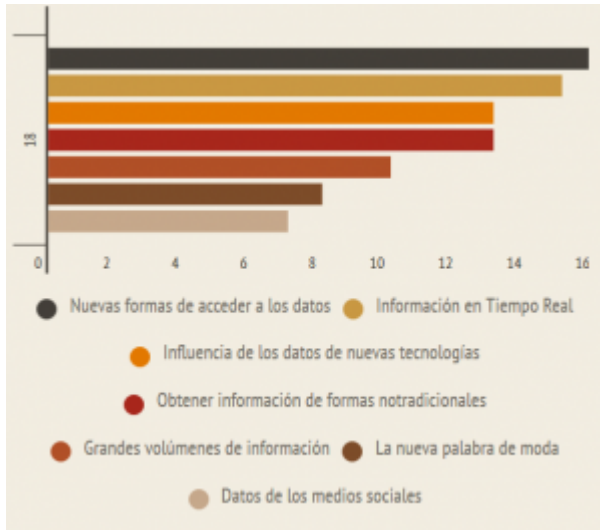


Imagen 1: Opinión de los directivos sobre qué es el big data

Una vez que se ha logrado obtener las diferentes opiniones, es el momento de definir lo que es el “big data”. Big Data es el tratamiento informatizado de grandes cantidades de información, la definición de lo que es Big Data no ha cambiado con el tiempo, puesto que los sistemas informáticos son cada vez más potentes y cada vez pueden almacenar y procesar más datos de lo que se podía antes. Además dependen de la capacidad del procesador, para algunos el problema está en procesar cientos de gigabytes, mientras que para otros se trata de petabytes cuando se encuentran con problemas.

“Big data” es un término aplicado a conjuntos de datos que superan la capacidad del software habitual para ser capturados, gestionados y procesados en un tiempo razonable. Los tamaños del “big data” se encuentran constantemente en movimiento creciente, de esta forma en 2012 se encontraba dimensionada en un tamaño de una docena de terabytes hasta varios petabytes de datos en un único data set. Debido a esto los métodos tradicionales para analizar y tratar toda esta información son poco útiles y se han ido explorando diferentes soluciones que lo que buscan es paralelizar el tratamiento de esta información.

Los datos que se analizan usando diferentes técnicas de “Big Data” pueden provenir de diferentes fuentes como son las siguientes:

- Generados por personas: a través del empleo de diferentes aplicaciones lo cual, genera metadatos que pueden ser analizados.
- Transacciones de datos: La facturación, las llamadas o las transacción entre cuentas generan información que tratada pueden ser datos relevantes. Este procedimiento también genera metadatos.

- La web: cada vez que una persona visita una web deja un rastro tras ella y a través de diferentes herramientas de tracking se podrán predecir sus futuros movimientos
- Machine to Machine: estos datos se producen al ser compartidos de un dispositivo a otro.
- Biométricos: son el conjunto de datos que provienen de la seguridad, defensa y servicios de inteligencia

Tras obtener estos datos, éstos se deben almacenar en único lugar, en un una base de datos no relacional, ya que las bases de datos tradicionales son incapaces de procesar cantidades tan grandes de datos.

Big Data, ¿es realmente algo nuevo?

Big data is like teenage sex: everyone talks about it, nobody really knows how to do it, everyone thinks everyone else is doing it, so everyone claims they are doing it

Con esta frase definió Dan Ariely lo que es el Big Data. Estas dos palabras se han convertido en una «buzzword». Actualmente todas las empresas dicen que están haciendo uso de estas nuevas tecnologías para analizar el mercado. Sin embargo, ¿es el Big Data algo nuevo, o lleva años ya entre nosotros? Hay quien dice que actualmente estamos viviendo la segunda era de Big Data. La primera era data en el siglo XIX entre los años 1820 y 1840. Era una era diferente pero muy similar a lo que vivimos hoy en día. A esta era Richards la denominó como la primera era de la información.



El siglo XIX era una época pre-digital diferente a la que nosotros conocemos. Los ordenadores no existían y en su lugar había personas dentro de despachos. En esta época ya se empezó a categorizar y clasificar la información. Esta información categorizada se disponía en numerosos sitios como librerías, bibliotecas, museos, etc. Además, los sistemas de catalogación se formalizaron. El almacenamiento y recuperación de datos se volvió un problema, por lo que se empezaron a crear bibliotecas científicas. Un ejemplo del problema de datos existente en aquella época se puede observar en los Blue Books de Reino Unido. Los Blue books son almanques u otra compilaciones de estadísticas e información. Estos libros se producían a un ritmo muy alto.

Por este motivo no eran capaces de procesar la gran cantidad de datos que se producían. Es por ello que William Playfair y Florence Nightingale crearon nuevas técnicas para visualizar estos nuevos datos.

Por otro lado, muchos de los datos que se recogían por los sistemas burocráticos se obtenían para tener el «control»: conocer a las otras clases sociales y ver si suponían una amenaza para las élites. De esta forma se podían observar los diferentes tendencias en la sociedad.

La información no es algo nuevo ni tampoco los datos. Actualmente, vivimos en un período el cual es llamado la segunda era del «big data» debido a la velocidad con la que se generan nuevos datos y la gran densidad de datos generados. Al igual que ocurrió con la primera revolución de los datos, en esta nueva revolución hay numerosas personas interesadas con los nuevos medios que están apareciendo. Las respuestas a esta revolución eran normalmente violentas, pero hoy en día tenemos mejores métodos para mantener el orden social.

Finalmente, me gustaría concluir con unas palabras de Patil:

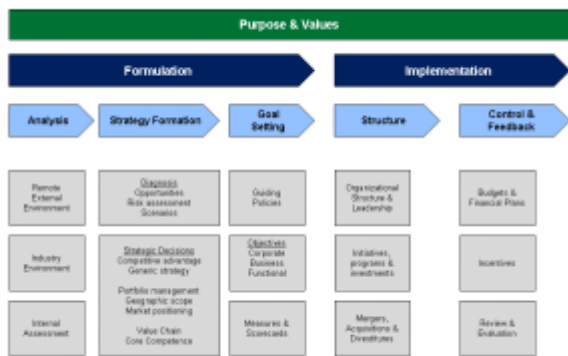
El reto del Big Data es saber utilizarlo de forma inteligente. Estamos en un momento en que todo el mundo habla sobre Big Data, pero igual que ocurre con el sexo entre adolescentes, todos hablan y nadie sabe cómo hacerlo. Por eso, los científicos de datos son quienes deben arrojar algo de luz sobre esta oscuridad.

[La planificación estratégica: ¿un asunto sólo de las empresas?](#)

Según la Wikipedia la planificación estratégica puede definirse de la siguiente manera:

La Planificación estratégica es un proceso sistemático de desarrollo e implementación de planes para alcanzar propósitos u objetivos. La planificación estratégica, se aplica sobre todo en los asuntos militares (donde se llamaría estrategia militar), y en actividades de negocios. Dentro de los negocios se usa para proporcionar una dirección general a una compañía (llamada Estrategia empresarial) en estrategias financieras, estrategias de desarrollo de recursos humanos u organizativas, en desarrollos de tecnología de la información y crear estrategias de marketing para enumerar tan sólo algunas aplicaciones.

Strategic Management Framework



Según esta definición, la planificación estratégica se utiliza esencialmente en el ámbito laboral de las empresas. Sin embargo, ¿es posible aplicarlo a otros ámbitos de nuestra vida cotidiana? La respuesta es afirmativa.

Hace unas semanas tuvimos la suerte de analizar nuestras vidas y ponerlo en común con otros compañeros de clase. En este ejercicio debíamos decir cómo nos veíamos dentro de cuatro años, es decir, elaborar nuestro plan estratégico de acuerdo a unos valores, misión y objetivos estratégicos.

Hacer esto es algo complicado, especialmente cuando ni tan siquiera eres capaz de imaginarte lo que harás mañana. El ejercicio estaba planteado y la respuesta al ejercicio parecía que no llegaría jamás. Pese a mis primeras impresiones, decidí comenzar el ejercicio cuanto antes, y así empezar a pensar cuanto antes lo que me gustaría que sucediese en un futuro lejano. Pronto me di cuenta de que escribir sólo mis objetivos es algo demasiado bonito. Había que concretar cómo haría que estos objetivos se convirtieran en realidad: cómo llegaría a alcanzar la meta que yo mismo me había propuesto conseguir. Hasta la fecha había logrado «planear» lo que haría en un espacio de tiempo medio. Una vez que había realizado esos planes, según pasaba el tiempo veía como muchos de los objetivos marcados desaparecían por falta de tiempo. Entonces, ¿cómo planear algo de dentro de cuatro años? Complicado, ¿verdad?. Poco a poco me di cuenta que no podía decir con total exactitud cómo resolvería los diferentes objetivos, pero sí podía dar algunas pinceladas de cómo lo haría y definir mis valores, misión y visión. Cuando todo esto estaba «decidido» se realizó una pequeña reunión con los demás compañeros. Al celebrarse esta reunión quedó algo claro: cuando empezamos la actividad, ninguno sabíamos muy bien cómo haríamos las cosas, pero sabíamos a donde queríamos llegar.

Tras la realización del ejercicio hay algo que me ha quedado claro: no se puede planificar todo. Esto no quiere decir que no deba planificar, sino que las planificaciones que haga deben ser flexibles y no cerradas. Además, todas ellas deben ir acompañadas de un colchón de tiempo por si algo sale mal. Acompañada de esta planificación se debe intentar ver posibles riesgos que puedan suceder para así encontrar una solución lo antes posible.