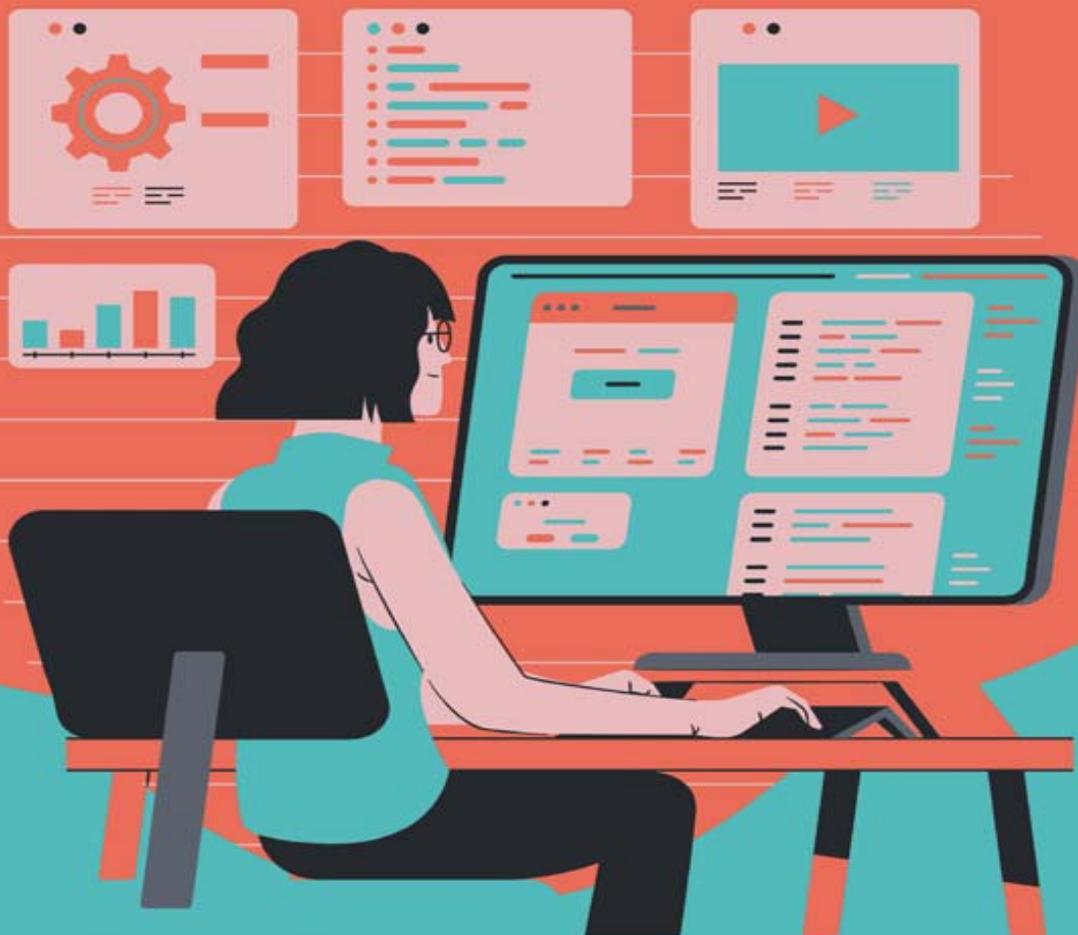




Libros.com

# TODO CUENTA

Introducción a la analítica de datos desde un enfoque metodológico



**JUSTO HIDALGO**

Una iniciativa de Tetuan Valley  
Prólogo de Camila Polensvaig

Primera edición digital: noviembre 2020  
Campaña de crowdfunding: equipo de Libros.com  
Composición de la cubierta: Patricia Á. Casal  
Maquetación: Álvaro López  
Corrección: Juan F. Gordo  
Revisión: Lucía Triviño

Versión digital realizada por Libros.com

© 2020 Justo Hidalgo

© 2020 [Libros.com](http://Libros.com)

[editorial@libros.com](mailto:editorial@libros.com)

ISBN digital: 978-84-18527-23-4



# **Justo Hidalgo**

## Todo cuenta

**Introducción a la analítica de datos desde un enfoque metodológico**

**Prólogo de Camila Polensvaig, CEO de Tetuan Valley**

*A Rosa y Olivia. Las dos chicas de mi vida, que me recuerdan cada día que todo cuenta, pero que no todo es medible.*

# Índice

[Portada](#)

[Créditos](#)

[Título y autor](#)

[Dedicatoria](#)

[Sobre Tetuan Valley](#)

[Prólogo](#)

[Introducción](#)

[1. En ocasiones veo datos](#)

[1.1. Fórmula 1](#)

[1.2. Datos comportamentales](#)

[1.3. Pulseras inteligentes](#)

[1.4. Termostatos](#)

[1.5. Indagando en el cerebro](#)

[1.6. Netflix y el contenido](#)

[1.7. Denodo y el acceso a los datos](#)

[1.8. Tesla](#)

[2. El modelo AARRR](#)

[3. Fase de adquisición](#)

[3.1. Observación](#)

[3.2. Promoción](#)

[3.3. Métricas de vanidad](#)

[3.4. Métricas que importan](#)

[3.5. El problema de la atribución](#)

[3.6. Seguridad y privacidad](#)

[3.7. Herramientas](#)

[3.8. El caso Cubelizer. Medir para mejorar también es posible en los espacios físicos. Por Celso Masid, co-CEO de Cubelizer](#)

[4. Fase de activación](#)

[4.1. A/B Testing](#)

[4.2. Distribuidores](#)

[4.3. Métricas que importan](#)

[4.4. Herramientas](#)

[4.5. El caso Datary. ¿Cómo se le ocurre a alguien intentar catalogar y centralizar las bases de datos de internet? Por Daniel Jadraque, CEO de Datary](#)

[5. Fase de retención](#)

[5.1. La formación de hábitos](#)

[5.2. Métodos para la retención de usuarios](#)

[5.3. Métricas que importan](#)

[5.4. Herramientas](#)

[5.5. El caso Teltoo. El análisis de datos como motor de I+D. Por Sergio Díaz-Miguel, CPO de Teltoo](#)

[6. Fase de referencia](#)

[6.1. Ejemplos de crecimiento referenciado](#)

[6.2. Métricas que importan](#)

[6.3. Herramientas](#)

[6.4. El caso Zensei. Cómo utiliza los datos y la analítica para ayudar a respirar mejor. Por David Martín-Corral, CEO de Zensei](#)

[7. Fase de ingresos](#)

[7.1. Introducción a los informes financieros básicos](#)

[7.2. Métricas para servicios de suscripción](#)

[7.3. Métricas para mercados](#)

[7.4. Herramientas](#)

[7.5. El caso Routive. Utilizando OKR en pequeñas organizaciones. Por Gabriel Domínguez, CEO de Routive](#)

[8. Breve introducción a las arquitecturas de datos](#)

[8.1. ¿Construir o integrar? La respuesta es «sí»](#)

[8.2. ¿Qué es un stack tecnológico?](#)

[8.3. Arquitecturas básicas de datos](#)

[8.4. El caso Kinequo. Cómo crear un producto de datos. Por Alberto Gimeno, CEO de Kinequo](#)

[9. Estrategia de analítica de datos en pequeñas organizaciones](#)

[9.1. Elaboración de una estrategia de datos](#)

[«La ética nos importa»](#)

[Conclusiones](#)

[Agradecimientos](#)

[Mecenas](#)

[Contraportada](#)

# Sobre Tetuan Valley

Tetuan Valley<sup>[1]</sup> es una organización sin ánimo de lucro que ofrece una escuela de *startups* para proyectos emergentes y nuevos emprendedores. Totalmente gratuita y sin ningún tipo de coste para los emprendedores, solo el máximo compromiso durante esas seis semanas. Fue en 2009 cuando Abel Muiño, compañero de trabajo en una vida pasada, me presentó a dos de los fundadores, Luis Rivera y Álex Barrera, para que les ayudase a conseguir una charla en la Universidad Nebrija, donde soy profesor asociado desde hace años, para captar equipos en su segunda edición.

Por curiosidad, empecé a asistir a algunas de las charlas como oyente, a conocer a algunos de los equipos, y poco después me pidieron que diese una de las charlas que se imparten durante las seis semanas de curso. Comencé con una de tipo «cajón de sastre» en la que, mediante un esquema de «Elige tu propia aventura», dejaba que los propios alumnos eligiesen de qué quería que hablase.

Desde el 2013 cambié el tema y me centré en la analítica de datos. Parece que gusta, porque me siguen invitando y cuando llego, la dirección que me han dado es la correcta. Aunque es una charla densa, intento que englobe lo que creo que hay que contar cuando se es mentor de equipos nuevos: ir al

grano, hablar de la teoría, pero con muchos ejemplos y referencias, y dejar claro a los equipos que, al final, la decisión acerca de lo que se mide y cómo, es suya.

Uno no se da cuenta del impacto que tienen iniciativas como Tetuan Valley hasta que no para un momento y reflexiona. Estamos hablando de decenas de equipos, cientos de potenciales emprendedores, que han pasado por las charlas, los *hackathons*, las sesiones de *mentoring*, las pizzas... No todos han seguido con sus proyectos. Pero eso también es bueno, ya que ha servido para que mucha gente se dé cuenta de lo complicado que es lanzar un proyecto. Otros han cambiado de proyecto. Otros han terminado trabajando para otras *startups*. No importa, pues el impacto positivo ya está hecho. El ecosistema existe, la profesionalización ya ha comenzado.

Que parte de mi tiempo pro bono lo dedique a Tetuan Valley es un placer. Pero hace años, Carmen Bermejo, en ese momento presidenta de Tetuan Valley, me ofreció ser parte del patronato de la organización. Un honor. Pero también una responsabilidad. Desde entonces, saco el poco tiempo que tengo para apoyar más a esta comunidad, sobre todo en mi labor de mentor, para ayudar a los emprendedores que empiezan a comprender algunos de los desafíos que van a encontrar. Nunca hay tiempo suficiente, y cuando nos juntamos en reuniones de patronato o durante alguna de las sesiones de Tetuan Valley, tengo la sensación de que tendría que darles más.

Este libro pretende continuar con ese apoyo a la comunidad de Tetuan Valley. Mi charla de «métricas», aumentada, corregida y revisada, se convierte en este libro que pretende introducir al nuevo emprendedor, al nuevo intraemprendedor y también al no tan nuevo, al apasionante mundo de la analítica de datos. Espero que como lector te sea útil y que, tras leerlo, tengas una idea mucho más clara de cómo los datos pueden ayudarte a mejorar tu producto o servicio.

El libro es de Tetuan Valley. Todos los beneficios obtenidos por la venta del libro irán a parar a Tetuan Valley, que legalmente es una asociación sin ánimo de lucro, para ayudarles a seguir adelante con programas formativos que ya han saltado nuestras fronteras.

# Prólogo

**Por Camila Polensvaig, CEO de Tetuan Valley (2017-2020)**

Tengo el enorme placer de formar parte de este proyecto junto a Justo Hidalgo y a otros grandes emprendedores, quienes cuentan, a través de su experiencia, cómo tomar las mejores decisiones posibles basadas en datos.

*Todo cuenta* es la suma de cientos de horas trabajadas, errores cometidos, decisiones acertadas y experiencias reales de diferentes emprendedores que enseñarán al lector a trabajar con la información y el análisis que necesita para su proyecto.

Aunque este libro es mucho más que eso.

Es la historia de diferentes personas maravillosas que llevan años trabajando en hacer realidad sus proyectos. Relatos de aquellos que han creado su *startup* desde cero, han cometido errores, aprendido de los mismos y han seguido trabajando hasta ver sus proyectos crecer y convertirse en lo que son hoy.

Este libro está formado por miembros de la Comunidad de Tetuan Valley, quienes han apoyado de forma incondicional a cientos de emprendedores y a quienes conoceréis a fondo a través de sus testimonios. Celso, Dani, Sergio, David, Gabriel y Alberto son parte de esta gran familia.

Le doy las gracias a Justo, a quien tanto admiro, por acercar estas historias a tantísimas personas. Por hacer realidad sus sueños y trabajar duro para ayudar

a los demás. Es una suerte haber hecho camino juntos, y qué felicidad que personas que no hayan podido hacerlo te conozcan y aprendan de ti a través de *Todo cuenta*.

Espero que este libro sea el primero de muchos. Porque dentro de Tetuan Valley hay muchas personas con historias que contar. Y muchos que estamos deseando leerlas y seguir aprendiendo de ellas.

# Introducción

Cuando una pequeña empresa empieza a pensar en la necesidad de analizar sus datos, suele hacerlo en términos de sus procesos financieros o de analítica web. En el primer caso, se centra en elementos como el flujo de caja, el balance, previsiones a varios meses o años, etc. En el caso de la analítica web, y tras unas búsquedas rápidas en el buscador de turno, encontrará una serie de herramientas para instalar en su sitio web y pasar a tener cierta información de visitas.

Este libro hablará de estos temas, pero el análisis de datos no acaba con la analítica web. Conocer fundamentos de estadística, de matemática discreta o de aprendizaje automático es cada vez más habitual en los negocios digitales —es decir, que son digitales per se o que utilizan lo digital para parte de su cadena de valor—. Ya no hay excusa para dejar de comprender el detalle del comportamiento explícito e implícito de los usuarios, para optimizar los procesos internos de la compañía y, en definitiva, para tomar las mejores decisiones a partir de la información y análisis producidas por los miles o millones de eventos que se realizan sobre nuestros servicios cada día. Lo que hace poco era solo posible para grandes corporaciones, se convierte en una

realidad y, por tanto, una obligación para cualquier empresa por pequeña que sea.

En este libro se desglosa la analítica de datos *lean* o ligera. Se describirán diversas metodologías, técnicas y herramientas, además de las métricas más interesantes. Se utilizarán casos reales, algunos muy conocidos y otros no tanto. Se introducirán algunos de los conceptos y procesos básicos de la ciencia de datos y se expondrán los desafíos y oportunidades de negocio que todo esto conlleva.

Este libro está pensado y diseñado para emprendedores y profesionales implicados en proyectos digitales innovadores y altamente competitivos. No es necesario tener un perfil técnico para sacarle partido a la obra, aunque sí es recomendable que el lector tenga la capacidad de comprender conceptos técnicos relacionados con el negocio.

Comenzaremos con una introducción a qué significa medir y a los modelos de análisis de datos más utilizados por parte de pequeñas organizaciones. Después, utilizando una de las metodologías más utilizadas en la actualidad, iremos definiendo cada una de las etapas con ejemplos reales y las métricas más utilizadas. Espero que, al terminar, el lector tenga una visión muy completa de todas las áreas que afectan al crecimiento exitoso de cualquier proyecto escalable y las métricas que nos permiten monitorizarlo. Si has montado una empresa o tienes la intención de hacerlo, necesitas conocer cómo gestionar sus datos. Si estás al cargo de un producto o servicio digital en una gran empresa, también. Este libro te ayuda a hacerlo bien desde el primer día. Comencemos.

# 1. En ocasiones veo datos

## 1.1. Fórmula 1

No es ningún secreto que los bólidos que compiten en la Fórmula 1, rodando a doscientos y trescientos kilómetros por hora, son ordenadores andantes, aunque no corrientes. Un automóvil estándar en Fórmula 1 puede tener hasta dos mil sensores que están obteniendo datos de todo tipo: la velocidad, pero también la presión de los neumáticos, el tiempo de trazado de curva y, por supuesto, todas las acciones que tiene un conductor de Fórmula 1 con los más de veinte mandos que tiene a su disposición. Estos sensores se utilizan para las simulaciones y pruebas antes de cada carrera, pero también durante la misma, donde los ingenieros de *boxes* obtienen estos datos en tiempo real. Aunque la telemetría existe desde los años ochenta, es quizá durante el Gran Premio de Brasil del año 2012 cuando el valor de los datos se demuestra de manera más clara. Sebastian Vettel estaba a punto de ganar su tercer campeonato del mundo. Tenía trece puntos de ventaja sobre Fernando Alonso, pero su bólido, tras un choque en la cuarta vuelta, no puede competir. Aunque no le impide seguir conduciendo, su velocidad y fiabilidad está muy por debajo del resto de competidores. Su gran rival, Fernando

Alonso, se escapa y parece que la suerte está echada. Lo que ocurre entonces es que los ingenieros de *boxes* se «ganan el sueldo»: durante las primeras diez vueltas de carrera, antes de que Sebastian Vettel llegue al *pit stop*, se reciben todos los datos del automóvil. Los ingenieros, a toda velocidad, simulan todas las posibles condiciones de carrera para que, cuando llegue el automóvil, elijan los ajustes necesarios al vehículo para que pudiese aguantar setenta vueltas más y, además, conseguir los puntos necesarios para poder ganar el campeonato del mundo. Y así ocurre. Sebastian Vettel queda sexto, obteniendo ocho puntos, más que suficientes para su objetivo. Aunque Fernando Alonso consigue ser segundo en la carrera, no es suficiente.

## 1.2. Datos comportamentales

De manera más general, un uso muy importante de los datos en la actualidad se basa en la comprensión del comportamiento del usuario al utilizar nuestra aplicación móvil o nuestro sitio web. Un caso típico es aquel en el que primero tenemos visitas en nuestra web —o descargas en nuestra app— que nos indica cierto interés. Posteriormente, estas visitas se convierten en usuarios registrados de manera gratuita, donde se les permite un cierto número de actividades que dependen de nuestro modelo de negocio. Después, un porcentaje de esos usuarios registrados se convierten en usuarios de pago. Solo saber que tenemos cinco mil, diez mil, cien mil, un millón o veinte millones de usuarios no es suficiente cuando queremos convencer a un inversor o cuando queremos negociar una colaboración con un posible socio industrial. Necesitamos conocer mucho mejor cuál es el comportamiento de los usuarios: qué perfil de usuarios se suele registrar, qué áreas de la web suelen visitar, qué contenidos suelen consumir dentro de nuestra aplicación... Todas son preguntas relevantes que tendremos que ser capaces de contestar a través de las métricas y del análisis posterior.

Contar con datos sobre el comportamiento ayuda también a mejorar el servicio, el producto. Podemos, con todo el cariño del mundo, pensar que hay una cierta funcionalidad que va a enamorar a nuestros clientes. Pero solo si contamos con datos cuantitativos del comportamiento podemos saber si realmente les está encantando esa funcionalidad o no de manera objetiva. Es

decir, si están entrando en esa funcionalidad, si la están utilizando y, por tanto si parece que es un éxito; o si entran, pero tampoco es algo que le está gustando demasiado; o si, por el contrario, nadie está utilizando esa funcionalidad. Con la información que nos dan los datos podemos inferir si no están entrando porque es una funcionalidad no deseada o no están entrando porque a lo mejor hemos diseñado mal la experiencia de usuario: el usuario no está encontrando esta funcionalidad.

A veces la importancia del uso de los datos viene obligada por las relaciones comerciales. Por ejemplo, cuando una tienda electrónica ofrece contenido de terceros (ya sea productos físicos o digitales), los propietarios de contenidos suelen exigir informes bastante detallados de cómo se están usando sus contenidos. En general, suele ser un informe de ventas —cuántas unidades se han vendido, a qué precio, en qué regiones—, pero en muchos casos la exigencia es mucho mayor. Nuestra organización tiene que ser capaz de obtener y procesar los datos adecuados para poder devolver esos informes en tiempo y forma.

### **1.3. Pulseras inteligentes**

Cada vez hay más empresas que basan su negocio en la comprensión absoluta de los datos propios o de los del usuario. Por ejemplo, empresas como Fitbit, que venden pulseras y otros dispositivos de medición de información corporal, desde el número de pasos realizados a lo largo del día hasta el ritmo cardíaco. La pulsera obtiene esos datos y después los sincroniza con una aplicación móvil, siendo capaz de darnos sugerencias tras un análisis posterior. Por ejemplo, quizás un usuario no está durmiendo bien con respecto a su perfil demográfico de edad y región. Otro usuario puede haberse puesto un objetivo de diez mil pasos al día pero lleva varios días andando apenas tres mil, por lo que se le envía una notificación para que «se ponga las pilas».

### **1.4. Termostatos**

Este interés en saber cómo se encuentra nuestro cuerpo se puede llevar también a nuestro hogar. Nest es una empresa que fue adquirida por Google en 2014 por tres mil doscientos millones de dólares norteamericanos. Podríamos pensar que esta empresa realiza ingeniería aeronáutica, pero no. Fabrica termostatos. Sí, esos dispositivos de toda la vida que tenemos en nuestra casa para saber y controlar qué temperatura hace. La única diferencia es que fabrican termostatos un poquito más inteligentes que los que compramos ahora mismo en cualquier ferretería. Se comunican a través de internet, con lo cual, se puede cambiar la temperatura desde un dispositivo móvil antes de llegar a casa. Pero además, también aprende del comportamiento. Nest va almacenando la información de uso y, a partir de ahí, pueden llegar incluso a proponer acciones concretas sobre la temperatura. Parece que por fin llegamos a la promesa de la domótica, el hogar inteligente que lleva anunciándose desde hace veinte años.

### **1.5. Indagando en el cerebro**

En el mundo de la creatividad existe un concepto que define el mejor producto como aquel que no existe, pero que cumple con todas las funciones que se esperan de él. Es el principio de idealidad. Así, la mejor correa para perros del mundo sería aquella que, sencillamente, no existiese, pero que aun así permitiese tener controlada y segura a nuestra mascota.

Las comunicaciones inalámbricas nos acercan en muchos casos a este objetivo utópico, pero siempre atractivo. Algunas aplicaciones y servicios dan un paso más, como Digits, gestionando los ahorros bancarios de una manera sencilla y sin una interfaz concreta, con una comunicación a base de mensajes de texto. Los relojes *smart* y las pulseras electrónicas también cumplen con parte de esa función. Un simple movimiento de muñeca me informa sobre las llamadas entrantes, las pulsaciones o el correo inesperado.

Un paso más está a punto de producirse, con la irrupción comercial de sistemas BCI, o *brain-computer interface*. Aunque con años de recorrido en el ámbito científico y clínico, y con empresas que ya llevan años ofreciendo este tipo de productos al mercado, es ahora cuando empiezan a surgir versiones a precio ajustado y capacidades que se acercan al mercado de masas. Sin ser

todavía una categoría de producto como las anteriormente mencionadas — relojes, móviles...—, el potencial es, en mi opinión, aún mayor, aunque también con grandes desafíos.



**Casco BCI Emotiv Insight.**

Un sistema BCI es aquel que permite obtener información de nuestro cerebro, generalmente a través de electroencefalogramas (EEGs), que es transformada en algún tipo de acción concreta por parte de un dispositivo o aplicativo. Su uso principal ha sido el clínico, procurando capacidades perdidas o inexistentes en pacientes, como por ejemplo el control mental de prótesis.

Sin embargo, desde hace de unos años, empresas como Emotiv están creando cascos comerciales a bajo coste que pueden ser utilizados tanto por pequeños laboratorios clínicos como por empresas de entretenimiento.

Es en el ámbito comercial donde los sistemas BCI tienen más posibilidades de crecimiento —lo cual, indirectamente, influirá positivamente en los proyectos de investigación—. Los factores de adopción de esta categoría son:

- **Tamaño.** En la actualidad, los sistemas BCI no intrusivos —aquellos que no exigen cirugía— son cascos bastante ligeros, pero todavía muy

aparentes. Si parte del fracaso de Google Glass fue su aparatosidad, está claro que los BCI adoptables han de ser aún más pequeños, donde apenas los terminales de medición sean visibles.

- Incomprensión. Aunque los móviles y los relojes han sufrido y sufren ciertas reticencias por parte de usuarios potenciales, debido principalmente a temas relacionados con la privacidad y propiedad de los datos..., aquí estamos hablando de empresas que almacenan información cerebral. Aunque en el estadio actual las señales eléctricas almacenadas no proveen ninguna clase de información útil a esas empresas, y no tendría sentido comercial venderla a terceros, cualquier decisión a este respecto conforma la base comercial y legal de lo que pueda llegar después.
- Ética y moral, *hacking*. Como se ha mencionado anteriormente, los sistemas BCI actuales son básicamente lectores EEG. No hay ningún tipo de retroalimentación al cerebro (más allá del calor que puedan generar los cascos). Pero ya existen cascos TDCS (*transcranial direct current systems*, sistemas transcraneales de corriente directa) que emiten corriente eléctrica a zonas específicas del cerebro, por lo que no es descabellado pensar en un futuro cercano donde sí haya esa retroalimentación al cerebro. Evidentemente, habrá muchos impedimentos legales y éticos. Se abre la puerta al «*hacking cerebral*», ya sea a propósito (personas que modifiquen las señales eléctricas que se envían de vuelta al cerebro para vaya usted a saber qué oscuros objetivos), o, quizá peor al principio, por error de los productos creados (un *bug* podría enviar señales incorrectas). Este será, sin duda, el punto que ralentizará el desarrollo de los sistemas BCI.
- Aplicaciones. La parte positiva de estos sistemas es el conjunto potencial de aplicaciones que se están creando y que se podrán crear en un futuro cercano. No solo videojuegos controlados mentalmente, lo cual, unido a la realidad virtual que ya estamos vislumbrando, daría paso al videojuego total, sino aplicaciones de meditación, de recomendación de productos (donde se demostraría si el *neuromarketing* funciona realmente o no), y, lo que me resulta más interesante, la utilización de aplicaciones existentes a través de un nuevo canal, un nuevo API cerebral que

exigirá una profunda reflexión acerca de cómo crear experiencias de usuario cuando las interfaces gráficas ya no son elementos accionables, sino de reflexión, recomendación y convencimiento a mentes pensantes.

No estamos ahí todavía, pero la rapidez con la que están apareciendo cascos EEG, aplicaciones y juegos acelera este proceso. Incluso tenemos los ya mencionados cascos TDCS para, aparentemente, ayudar a mejorar la concentración, etc.

## **1.6. Netflix y el contenido**

No podemos hablar de análisis de datos sin mencionar a Netflix. Netflix es un servicio de suscripción. Pagando una cuota al mes se pueden ver miles de series y películas. Netflix comenzó su andadura como un servicio de alquiler de DVD en Estados Unidos de América. En lugar de tener que ir al videoclub a alquilar la película, se hacía —y se sigue haciendo— a través del correo ordinario. El usuario seleccionaba la lista de películas que quería ver y en qué orden, y el DVD de la primera película (o las primeras, dependiendo del tipo de suscripción que eligiese) llegaba a su casa. Cuando el usuario veía la película, la podía devolver en cualquier buzón y tras su recepción, Netflix enviaba la siguiente de la lista. Hacia el año 2007, Netflix comenzó a ofrecer su servicio de *streaming*. Ya no hace falta que llegue el DVD físico, sino que, con la conexión casera de internet, el usuario puede conectarse y ver la película en tiempo real.

Uno de los ejemplos paradigmáticos de utilización de datos para mejorar servicios es su sistema de recomendaciones. Netflix no tiene, ni tendrá jamás, a menos que cambien mucho las cosas, todas las películas del mundo. Por tanto, a Netflix no le interesa tanto lo que busque el usuario como lo que descubra. Por ello, su sistema de recomendación es parte intrínseca del motor que hace que la gente utilice Netflix casi cada día. Así, recomienda películas de acuerdo a los gustos del usuario, experiencias previas, películas o series que hayan visto personas relacionadas de una u otra manera, etc. En el año 2006, Netflix anunció un concurso abierto para mejorar su algoritmo de

recomendación, con un premio de un millón de dólares. Muchas empresas e instituciones académicas y de investigación se presentaron. El equipo «Bell Kor's Pragmatic Chaos», con componentes de empresas como Commendo Research & Consulting GmbH, AT&T Labs y Yahoo!, ganó el premio. Pero años después se supo que Netflix no llegó a utilizar las ideas del proyecto ganador. Es en ese momento cuando decidieron enfocarse en el servicio de *streaming*. Eso significa que los datos que se tienen de los usuarios cambian totalmente, lo cual implica que el modelo de predicción ganador del premio ya no servía. Ahora Netflix tiene muchísima más información de los propios usuarios. Antes solo se podía saber la película alquilada, la fecha de envío, la fecha de recepción y poco más. Ahora se puede saber cuándo se empieza a ver la película, si se termina o no, si se da a pausa, si los capítulos de una serie se ven de dos en dos o de tres en tres... Con esa información, Netflix se ha adaptado a nuestros nuevos hábitos de visualización de contenido, como el *binge-watching* —la visualización de series viendo varios, e incluso todos los capítulos de una temporada de una vez—, o el famoso lanzamiento de la adaptación de la serie británica *House of Cards* basándose en los datos de los usuarios de Netflix.

## 1.7. Denodo y el acceso a los datos

Como iremos viendo a lo largo del libro, según vayamos avanzando en nuestro negocio iremos requiriendo nuevos tipos de datos. Quizás al comienzo solo necesitaremos los datos provenientes de nuestra herramienta de analítica web. Más adelante querremos utilizar mapas de calor, realizar pruebas A/B, o añadiremos una aplicación móvil nativa de la que también querremos obtener información. Quizá necesitaremos acceder a información regulatoria, datos públicos de internet, etc. ¿Cómo conseguimos acceder, procesar y analizar tanta información proveniente de tantas fuentes de datos, tanto internas como externas?

La primera empresa en la que trabajé con proyectos de datos de alto calado fue Denodo. La que ahora es una de las empresas líderes en una categoría de producto llamada Virtualización de Datos, comenzó su andadura con proyectos de integración de información web para crear comparadores de

productos y agregadores financieros. En estos proyectos, creamos tecnología que permitía acceder a páginas web e interactuar con ellas de manera automática para obtener datos como si se tratasen de puras bases de datos. Poco a poco, Denodo fue añadiendo diferentes tipos de fuentes de datos, desde las típicas bases de datos relacionales a ficheros de texto, PDF, etc., pasando por la integración con herramientas comerciales como SAP. La característica principal del producto implementado es que no requería «copiar» los datos de la fuente original a una base de datos local —algo habitual con otros enfoques como el del almacén de datos, como veremos en la sección 12.3—. Al contrario, cuando el usuario necesita acceder a esos datos, Denodo accede directamente a las fuentes, permitiendo una gran «frescura», algo muy importante cuando estamos en mercados críticos, como el financiero o el de salud. Poco a poco, este tipo de producto ha evolucionado y añadido capacidades mucho más avanzadas, como cachés que permiten almacenar temporalmente datos procedentes de algunas fuentes que no requieren de esa frescura mencionada anteriormente, o funcionalidades de optimización de consultas para que los tiempos de respuesta sean mucho más rápidos[2].

## 1.8. Tesla

Obviamente, tampoco podemos dejar de mencionar a Tesla. La fábrica de automóviles creada por uno de los fundadores de PayPal, Elon Musk. Esta persona decidió hace ya muchos años que iba a crear el primer automóvil eléctrico comercial del mundo. Cuando se empezaron a vender estos coches, parecía solo para locos: precios altísimos y autonomía de cincuenta a cien kilómetros. Pero ya podemos ver en muchas partes del mundo cómo hay Teslas y otros modelos. Quizá todavía para un cierto nivel económico, pero con ventas crecientes y un interés público por eliminar los automóviles de gasolina y diésel. Pero lo importante de Tesla para este libro no es que sea un automóvil eléctrico, sino su otra característica principal: la idea de que sea un automóvil autónomo, es decir, autoconducido. Al igual que los Fórmula 1 tienen dos mil sensores y la capacidad de procesar datos en tiempo real, los automóviles Tesla tienen entre doce y veinte sensores que capturan la

información en tiempo real, tanto de lo que ocurre a nuestro alrededor como de su comportamiento a los mandos. Cuando un Tesla llega al mercado, ha sido entrenado previamente con cientos de miles de vídeos que informan al coche de qué es lo que ocurre y cómo tiene que actuar. Por ejemplo, si aparece una pelota botando por delante, el automóvil aprende que lo más probable es que después aparezca un niño corriendo detrás de ella, por lo que no ha de arrancar inmediatamente. Todavía no nos encontramos en el estadio en el que el anteriormente denominado conductor pueda tumbarse en la parte de atrás a echarse una siesta mientras el automóvil le lleva a su destino, pero sí hay un número incremental de acciones que no hay que controlar manualmente.

## 2. El modelo AARRR

Como podemos ver incluso en la prensa generalista, hay cada vez más interés y capital en cómo las empresas pueden coger datos del entorno de sus productos y servicios y tomar decisiones que hace apenas diez años solo estaban al alcance de grandes conglomerados. Estos temas no solo son ahora interesantes, sino «sexys», tal y como definió la revista *Harvard Business Review* al trabajo de científico de datos.

Y sin embargo, esto no ha llegado a todo el mundo. Sigue habiendo un porcentaje no trivial de emprendedores que se lanzan a la aventura digital sin un mínimo conocimiento de cómo medir sus resultados. Como mucho, habrán oído hablar de Google Analytics, una herramienta de Google con una versión gratuita que sirve para medir el comportamiento de las visitas a nuestro sitio web[3]. Analytics es una gran herramienta para medir el comportamiento de los usuarios en las páginas web; es una herramienta muy buena para saber cuántas visitas tenemos; de dónde vienen; cuál es la demografía de nuestros usuarios. Pero no es en absoluto suficiente para el mundo en que nos encontramos hoy. Primero, porque hay muchas más herramientas para muchas otras actividades relacionadas con la medición y el análisis de datos. Pero además, porque antes de ponernos a instalar Google