MOOC Internet de las cosas IOT

MOOC IOT

Módulo 6. IMPACTO DE IOT EN LOS NEGOCIOS

6.1 INTERNET DE LOS CONSUMIDORES

Por Juan Sánchez Fernández

Departamento de Comercialización e Investigación de Mercados. Universidad de Granada.

---

1. Introducción

Las múltiples vías de conexión empresa-consumidor que se han abierto en los últimos años gracias al desarrollo espectacular de las tecnologías de la información y comunicación han brindado la posibilidad de un conocimiento mucho más amplio y preciso de la base de clientes reales o potenciales por parte de las empresas. Esto abre la posibilidad de relaciones mucho más estrechas y estables con los mismos, en gran parte sustentado en una mejor adaptación de la oferta que conlleva una mayor satisfacción, así como en unos flujos de comunicación más directos y rápidos. A través de lo anterior, se refuerza el papel de *prosumer* que hace sentirse alos propios consumidores más valorados por la empresa. A las ya clásicas redes sociales (Facebook, Instagram, Twitter….) como canales de comunicación y vías de participación de los consumidores, se han unido en los últimos años las posibilidades que expande IoT, objeto principal de este MOOC.

Desde el punto de vista de la gestión del marketing, en su función principal de relación con los mercados, el hecho de tener dispositivos conectados de una manera permanente a una red, generando datos sobre uso, momentos de uso, frecuencia de uso, condiciones de uso, etc. de cada vez una mayor base de productos, abre la posibilidad de, entre otras cosas, hacer segmentaciones de consumidores mucho más precisas, diseñar productos mucho más adaptados a las necesidades reales de los consumidores, ofrecer servicios adicionales que generen un mayor valor añadido para los mimos, así como alimentar un CRM que permita unasrelaciones mucho más estables y duraderas empresa- consumidor.

En los próximos epígrafes se aborda cómo la obtención de una información más rápida, fiable y económica sobre el consumidor gracias IoT abre nuevas vías de relación empresa-cliente que hacen posible una mayor satisfacción y lealtad, objetivos fundamentales en la gestión del marketing.

UNIVERSIDAD DE GRANADA abiertaugr la formación abierta de la UGR CC BY NC ND

2. Conocimiento del consumidor en el contexto de IoT

El conocimiento del consumidor a través de la investigación de marketing ha dado un paso más gracias a la materia prima que se genera en el contexto de IoT. Efectivamente, se pasa de una experiencia de consumo fundamentalmente privada por parte del consumidor a una nueva situación donde se genera un incremento exponencial de datos cualitativos y cuantitativos de la experiencia de consumo gracias a los productos conectados. Para ser más preciso, los métodostradicionales de investigación de mercados, fundamentados en encuestas o entrevistas en profundidad (en grupo o individuales), suelen implicar la selección de una muestra más o menos representativa para obtener, vía cuestionarios o entrevistas, información de utilidad que permita conocer la opinión de los consumidores. En el nuevo contexto, se puede obtener información mucho más precisa y fiable no ya de una muestra de consumidores, sino de todos los consumidores, en cuestiones tan importantes para ofrecer un mejor servicio al cliente como cuándo, cómo, cuánto, dónde, en qué condiciones, en qué lugar, etc. se usa un producto. Todoello ofrece nuevos conocimientos para la creación de valor, contribuyendo a una nueva relación de servicio donde las nuevas experiencias son co-creadas [1]. Si se combinan esos datos con otras herramientas como geolocalización y bases de datos de carácter sociodemográfico, las posibilidades de la investigación de mercados crecen de manera exponencial [2]. Además, cualquier información adicional introducida por el propio usuario a través de herramientas como de las apps enriquece aún más esas bases de datos.

Un buen ejemplo de los datos que se pueden generar y compartir cliente/empresa lo encontramos en el gigante norteamericano Garmin, con su gama de productos para medir la actividad físico-deportiva. Suele ser habitual después de adquirir un dispositivo de la amplia gama que ofrece Garmin registrarse on-line para acceder a una serie de servicios adicionales (figura 1).

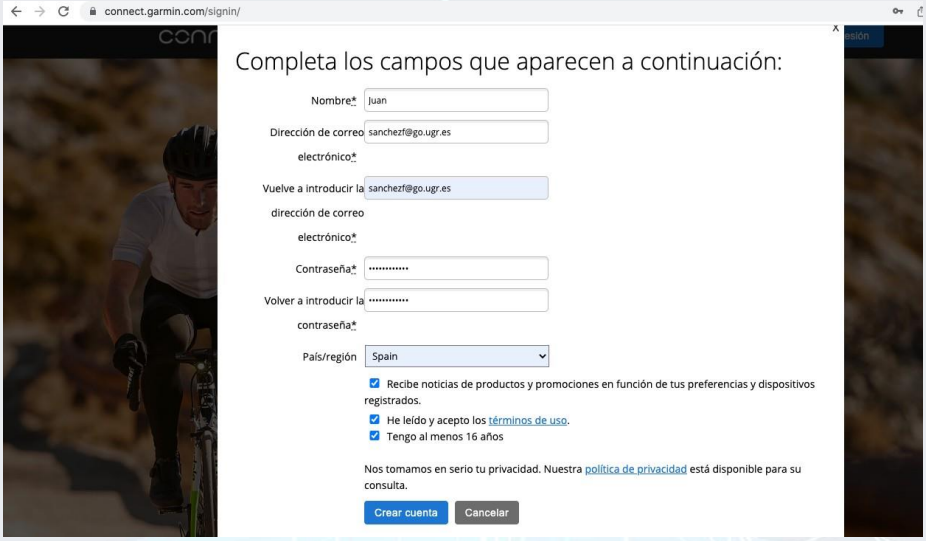

Figura 1. *Crear cuenta en Garmin Connect*

UNIVERSIDAD DE GRANADA
abiertaugr la formación abierta de la UGR
BY NC ND

En el propio proceso de registro, se pide información adicional del usuario para obtener todo el rendimiento posible a las funcionalidades añadidas que se ofrecen por parte de la empresa online (figura 2).

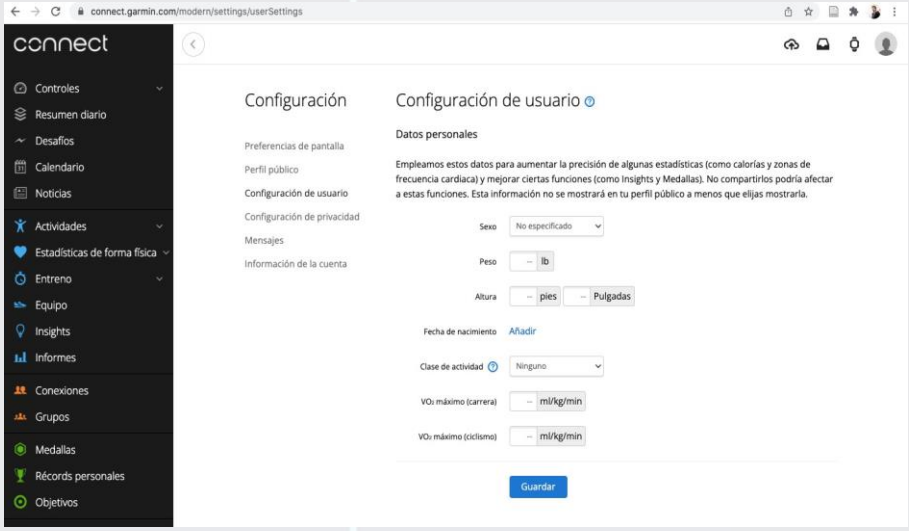

Figura 2. *Proceso de Registro en Garmin Connect*

A modo de ejemplo, el Garmin Forerunner Fenix es uno de los dispositivos más populares a nivel mundial en el ámbito del deportivo. Se trata de un dispositivo tremendamente completo que registra la actividad física de los usuarios, recopilando información sobre una amplia variedad de actividades deportivas (ver figura 3).

Figura 3. *Dispositivo Garmin Forerunner Fenix 6*

UNIVERSIDAD DE GRANADA
abiertaugr
la formación abierta de la UGR
BY NC ND

El dispositivo en sí almacena internamente gran cantidad de datos, pero las funcionalidades aumentan con sus posibilidades de conectividad y los análisis adicionales que es posible realizar cuando esos datos se transfieren a los servidores de Garmin gracias a una aplicación desarrollada por la empresa que se denomina Garmin Connect. Esta aplicación se puede instalar en un *smartphone* y ofrecer una información mucho más completa y con mayor riqueza visual.Además, suele actuar como una de las posibilidades de trasladar los datos desde el dispositivo a la base de datos de la empresa gracias a una conexión *bluetooth*, que actúa de nexo entre el dispositivo y el *smartphone*, para al mismo tiempo trasladarse desde el *smartphone* a la base de datos de Garmin a través de Internet.

Figura 4. *Garmin Connect app*

La información que se almacena en el teléfono móvil es muy rica, pero más aún lo es la que se concentra en la base de datos de Garmin. En cualquier caso, cliqueando en cualquier actividad de las desarrolladas por el usuario se obtiene información ampliada como, por ejemplo: distancia recorrida, tiempo empleado, calorías consumidas, tiempo medio, etc. En las figuras 5 y 6 podemos ver una muestra del tipo de información que registra Garmin después de cada actividad deportiva.

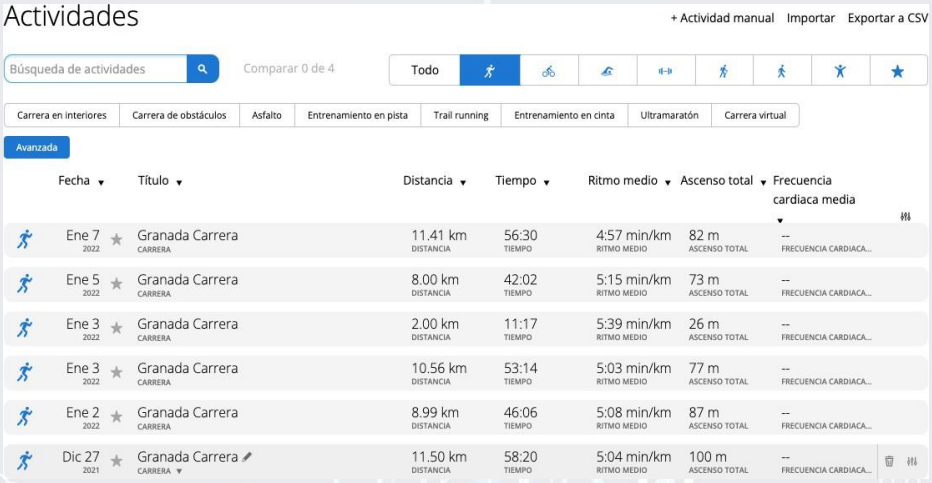

Figura 5. *Actividades registradas en Garmin Connect*

UNIVERSIDAD DE GRANADA
abiertaugr la formación abierta de la UGR
CC BY NC ND

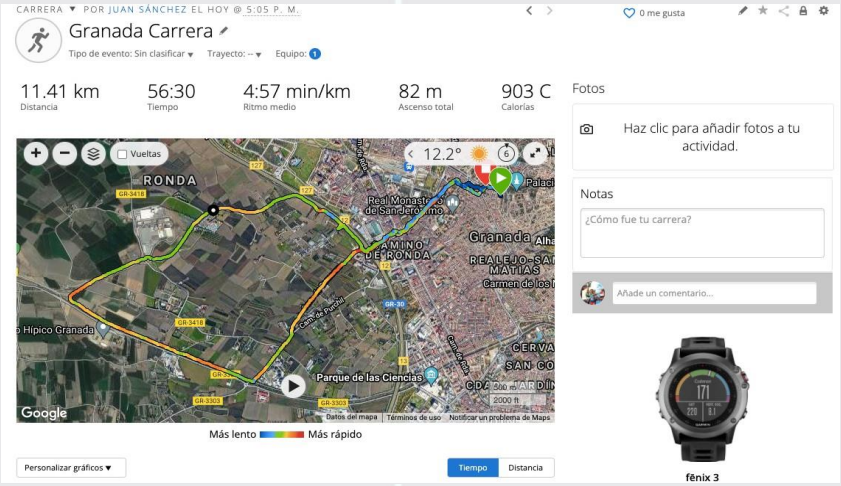

Figura 6. *Servicios ofrecidos por Garmin Connect a*

Además, esa información se puede detallar aún más, en forma de gráficos personalizables por el usuario como los que se pueden ver en la figura 7.

Figura 7. *Servicios ofrecidos por Garmin Connect b*

El propio usuario (figura 8) puede editar cada una de las actividades e incluir información adicional como por ejemplo, vídeos, fotos, comentarios sobre el recorrido, cómo se ha sentido durante la carrera, dispositivo empleado, material deportivo utilizado, etc..

UNIVERSIDAD DE GRANADA
abiertaugr
la formación abierta de la UGR
BY NC ND

Figura 8. *Información ampliable por el consumidor en Garmin Connect*

Por tanto, esa visibilidad de la experiencia de consumo cambia la unidad de análisis, lo que posibilita un conocimiento mucho más amplio donde los comportamientos no sólo se van a explicar en términos de **disposición** (características personales) sino también de **situación** (características contextuales). Está claro que los factores situacionales pueden llegar a ser tanto o más importantes para la satisfacción del consumidor. Efectivamente, IoT hace visibles las conexiones e interacciones en momento y lugar, prácticamente de cualquier producto o servicio a través de diferentes posibilidades de red, lo que podría revelar patrones reales de comportamientos para entender las motivaciones del consumidor de una manera nueva que noera posible anteriormente, al menos no tan precisa.

Además, los dispositivos conectados a IoT a través de una app instalada en un Smartphone quesuele ser suministrada por el fabricante son una vía excepcional de mejorar el CRM de la empresa. En este contexto, Saarijärvi et al., (2013) consideran que el CRM ha llegado a ser más importante a medida que las empresas paulatinamente cambian su atención de los bienes a los servicios y de la venta al servicio.

3. Nuevos productos/servicios

Ese continuo flujo de datos abre la puerta a una oferta que tiene el potencial de ser dinámica. Así, los productos tienen la posibilidad de adaptarse mucho mejor por ejemplo a través pequeñas aplicaciones informáticas que permiten una personalización parcial para dar respuesta a situaciones de uso específicas. Dicha de otra forma, las necesidades del cliente pueden ser satisfechas de una manera mucho más adecuada gracias a una combinación de características físicas y digitales de un producto, basadas estas últimas en los datos aportados, en la mayoría de los casos de forma automática, por el cliente. A partir de un producto más o menos estándar, es posible ir adaptando sus capacidades gracias a la adición de servicios digitales que respondan a situaciones concretas de los clientes [4]. Efectivamente, muchas empresas se están enfocando en la estandarización,

UNIVERSIDAD DE GRANADA
abiertaugr la formación abierta de la UGR
CC BY NC ND

arquitectura modular y la escalabilidad de las soluciones que desarrollan, a la vez que ofrecen a los consumidores la opción de customizar e integrar estos módulos de acuerdo con sus necesidades.

Esto desde el punto de vista del consumidor es especialmente importante, pues abre la vía de pago en función del uso real que se hace del producto/servicio. Por tanto, la nueva forma de satisfacer las necesidades de los consumidores con los productos conectados conlleva un cambio en los patrones de propiedad de los productos, donde la misma ya no recae fundamental y exclusivamente en manos de los clientes, sino que está será de la propia empresa, pagando los clientes en función del uso y condiciones de uso de los mismos (un claro ejemplo son las bicicletas de alquiler que hay en cada vez más ciudades, que no reducen la necesidad de comprar y almacenar una bicicleta, sino que también son un claro sustituto del uso del coche privado y otros sistemas de transporte público). Incluso es factible llegar a que los consumidores no paguen en dinero, sino en datos, por disfrutar de algunos servicios, como por ejemplo ocurre en caso de Google.

Esto es perfectamente posible porque las ofertas tienen nuevas capacidades que son dinámicas, incluso después de que un producto haya sido diseñado, fabricado y vendido. Por tanto, en este entorno, el servicio cobra mucho más valor, y deber ser realmente la base de la diferenciación que permita a las empresas obtener y mantener su ventaja competitiva. Tanto es así, que las empresas evolucionarán a entidades proveedoras de servicios [5].

Siguiendo con el ejemplo de Garmin, a través de Garmin IQ (figura 9), la empresa abre la puertaa otros desarrolladores que ofrecen funcionalidades para el dispositivo, unas de manera gratuita y otras de pago. Esto supone una fuente de servicios e ingresos adicionales para Garmin.

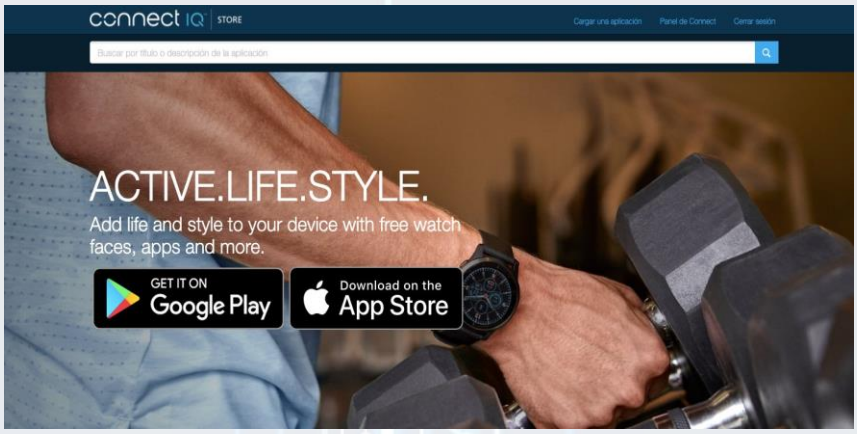

Figura 9. *Garmin iQ*

UNIVERSIDAD DE GRANADA
abiertaugr
la formación abierta de la UGR
BY NC ND

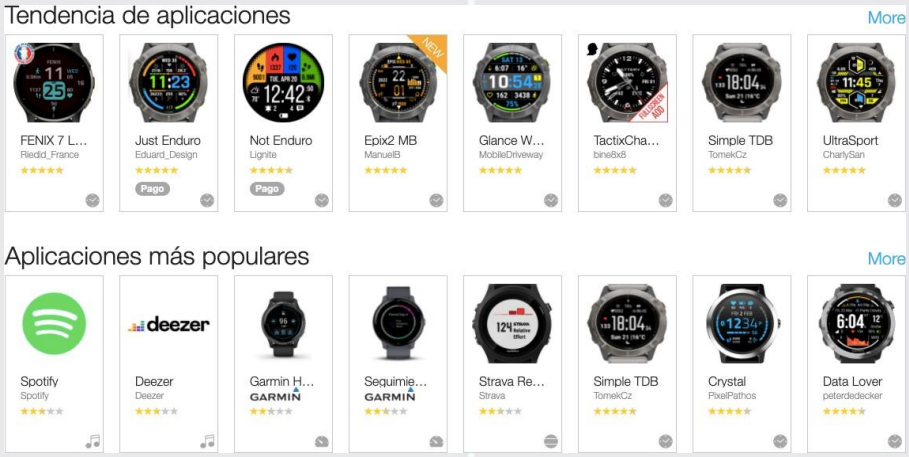

Figura 10. *Aplicaciones de Garmin IQ*

A esto algunos expertos lo denominan plataformas dinámicas de servicio [1]: Dinámicas, porque la tecnología material y digital permiten la reconfiguración y la movilidad y puede ser experimentado cada vez que el consumidor quiera; servicio porque se aplicarán las competencias para crear valor con los consumidores; plataforma, porque permite a los consumidores adecuar mejor su experiencia con el producto a los contextos específicos y adaptar sus funcionalidades a través de sus acciones.

Una de las funciones principales del marketing en este nuevo contexto será la generación de valores, la satisfacción de necesidades latentes, a la vez que el fomento de nuevos comportamientos de los consumidores, algunos de ellos relacionados con la co-creación a través de la intervención de diseño, que permite a los consumidores contribuir en la configuración de los productos inteligentes, lo que en última instancia también puede incrementar su propia satisfacción.

Efectivamente, las posibilidades de configurar el producto con características digitales debe serla base de una mayor satisfacción del cliente, preludio de una posterior fuerte lealtad a la empresa. En este sentido, no solo se puede adaptar el producto con servicios digitales, sino que también gracias a la información que se va obteniendo de las diferentes actividades registradaspor los propios usuarios, la empresa puede disponer de ideas para innovar en productos. De manera general, las posibilidades de desarrollar nuevos productos/servicios son diversas, habiendo al menos seis formas de generarlos:

- Nuevo producto para todo el mundo
- Nuevos productos para la empresa
- Nuevos productos en una línea de producto
- Mejoras y revisiones de productos existentes
- Reposicionamiento de los productos existentes
- Reducciones de precio a través innovación en diseño o procesos

UNIVERSIDAD DE GRANADA
abiertaugr la formación abierta de la UGR
CC BY NC ND

El éxito de un producto dependerá fundamentalmente del valor que genera para los consumidores a los que se destina, de forma que tendrá más posibilidades de éxito si incorpora la información del consumidor. Sin embargo, esto no es siempre conocido por la empresa, incluso los propios consumidores pueden no ser conocedores explícitos de las mismas, no obstante, su situación de consumo resumida a través de los datos que se generan gracias a IoT contribuye a solventar esa tradicional falta de conocimiento.

Esa misma información puede ser la base de una de las claves de la gestión del marketing en materia de producto como es caso de la **diferenciación**. No obstante, teniendo en cuenta que los contextos de consumo del consumidor son muy diversos, una oferta adaptada a esos múltiples contextos podría ser extremadamente costosa para la empresa. Sin embargo, los servicios complementarios podrían utilizarse para abordar la variedad, y por tanto adaptación y diferenciación, y los servicios básicos permitir a la empresa mantener cierta estandarización y lograr una mayor eficiencia en los costes.

Si los recursos de personalización son digitales en naturaleza, la microsegmentación se consigue teniendo en cuenta que la tecnología digital soporta la flexibilidad de diseño a lo largo de ciclo de vida del producto. Los *smartphones* y sus apps son un claro ejemplo de personalización y diferenciación del producto a través de la autoconfiguración y autoselección.

Con esa finalidad de personalización, los objetos conectados tienen una serie de característicasque los hacen especialmente adecuados [6]:

- *Sensibilidad*. Capacidad de un producto de sentir y responder a los cambios ensu entorno.
- *Direccionabilidad* y *traceabilidad*, en la medida que el objeto puede ser localizado en tiempo real.
- *Asociabilidad*, en la medida que puede ser asociado y coordinado con otros objetos, para hacer inferencias sobre futuros estados y condiciones.
- *Comunicabilidad*. Capacidad de enviar y recibir mensajes.
- *Programabilidad*. Capacidad el producto de aceptar nuevos conjuntos de instrucciones y modificar su comportamiento.
- *Memorabilidad*. Capacidad de almacenar información y registros históricos desu estado e interacciones.

Todo esto permite expandir la materialidad física de los productos a través de capacidades digitales basadas en software. Más concretamente, la inteligencia y la conectividad posibilita un nuevo conjunto de funcionalidades en los productos que pueden ser agrupadas en 4 áreas [7]:

UNIVERSIDAD DE GRANADA
abiertaugr la formación abierta de la UGR
CC BY NC ND

- Seguimiento. Los sensores y fuentes externas de datos permiten un seguimiento completo de:
    - Las condiciones del producto
    - El entorno externo
    - El funcionamiento y uso del producto

- Control. El software integrado en el producto o en la nube con la que se conectael producto hace posible:
    - Controlar las funciones del producto
    - Personalizar la experiencia de uso
- Optimización. El seguimiento y las capacidades de control posibilitan algoritmosque optimizan el funcionamiento del producto en orden a:
    - Mejorar el rendimiento del producto
    - Permitir diagnósticos predictivos de servicio y reparación
- Autonomía. La combinación de seguimiento, control y optimización permite:
    - Funcionamiento autónomo de los productos
    - Auto-coordinación de funcionamiento con otros productos y sistemas
    - Mejora y personalización autónoma del producto
    - Autodiagnóstico y servicio

Los productos autónomos también pueden interactuar con otros productos y sistemas. El valorde estas características aumentará a medida que más productos vayan teniendo las capacidadesde conexión.

4. Internet de las cosas como sistema de ensamblaje o servicio ampliado

El principio básico de todo esto es que la conceptualización de IoT como un todo implica que es más que la suma de las partes, y la identidad del todo se construye por sus componentes y las propias interacciones de sus componentes heterogéneos. En este sentido, la colaboración de diferentes empresas amplia de manera significativa el servicio que se disfruta. Para el ejemplo que venimos empleando, la posibilidad de compartir los datos de Garmin con aplicaciones como Strava o Relive añade valor al consumidor, en parte de manera gratuita, y en parte pagando adicionalmente por la misma.

UNIVERSIDAD DE GRANADA
abiertaugr la formación abierta de la UGR
CC BY NC ND

Figura 11. *Servicios complementarios a Garmin*

Strava es como una especie de Facebook de deportistas. No es necesario disponer de un dispositivo físico de Strava, sino que es completamente una empresa de servicio. Apoyándose en los registros de Garmin Connect (entre otras muchas otras posibilidades), Strava suministra un valor añadido adicional al usuario de *wereables* deportivos como Garmin.

El producto estrella de Strava desde su origen fueron los segmentos, que son tramos habituales de los usuarios con los que la aplicación realiza de manera automática realiza una clasificación precisa de los tiempos de los mejores deportistas, de indudable interés para los usuarios, pues son la base en muchos casos de una verdadera gamificación. Además, Strava es una herramienta muy potente de analítica entrenamientos, de especial interés para saber si se está o noprogresando, hacer comparaciones con otros momentos temporales, con otros usuarios conocidos o no, etc. En su versión de pago, se puede acceder además a unos servicios adicionales muy interesantes entre los que destacan [8]:

- Rutas, planificación y recomendaciones
- Combina paseos y carreras con análisis de rendimiento
- Registro de entrenamiento
- Tablas de clasificación de segmentos, leyendas locales
- Segmentos en vivo en dispositivos compatibles
- Tendencias y comparaciones de actividad mensual
- El establecimiento de metas
- Mapas de calor
- Planes de formación
- Curva de potencia: si tiene un medidor de potencia
- Análisis de fitness y frescura

Pero realmente, Strava no solo trabaja con datos de Garmin. Hoy en día, prácticamente no hay ningún dispositivo GPS de prestigio que no ofrezca la posibilidad de subir las actividades a Strava sin más que haberlo autorizado previamente por parte del usuario. Entre las más conocidas se encuentran: Amazfit, Bkool, Bryton, Casio, Fitbit, Garmin, Polar, Sigma, Soleus o Suunto. Hoy en día realmente es un *backup* de todas las actividades de un usuario de GPS deportivo, con independencia de que a lo largo de su trayectoria este haya cambiado de marca.

UNIVERSIDAD DE GRANADA
abiertaugr la formación abierta de la UGR
CC BY NC ND

Por su parte, Relive, alimentándose también de datos procedentes de dispositivos deportivos GPS, como Garmin, ofrece nuevos servicios, fundamentalmente relacionados con imagen,sonido y video. En definitiva, aporta valor partiendo de datos de dispositivos GPS de otras empresas.

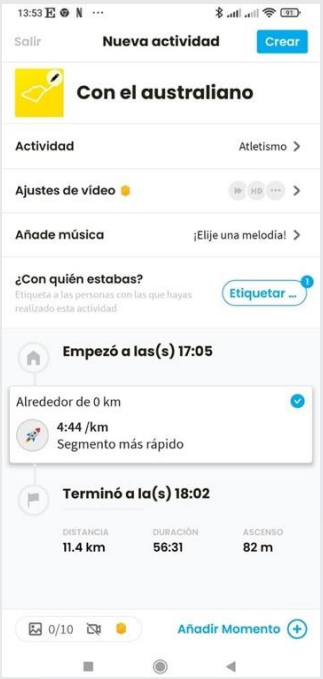

Figura 12. *Servicios complementarios Relive*

Pero es que además de estas relaciones entre empresas con un marcado carácter tecnológico, se abre la posibilidad de entrada a otras empresas que ofrezcan productos y servicios adaptados a las circunstancias especiales de los clientes, gracias a los datos que generados con algunos desus dispositivos conectados. En este sentido, siguiendo el ejemplo de Garmin, con sus dispositivos GPS conectados recaba información de interés para otras empresas de material deportivo (se dispone de información sobre el tipo de terreno por el que corre un deportista, la hora del día a la que suele correr, las veces que lo hace por semana, las distancia media recorrida, el ritmo por kilómetro, etc., potencialmente de interés para empresas como por ejemplo Adidas a la hora de ofrecerle a esos deportistas el tipo de calzado o ropa más indicada a sus circunstancias de uso, o estimar en base a la propia información introducida por el usuario en relación al calzado empleado, y la registrada por el dispositivo en cuanto a los kilómetros recorridos, cuándo es recomendable cambiar de calzado), de seguros (se recoge información como veces a la semana que se practica deporte, tiempo empleado en diferentes actividades deportivas o frecuencia cardíaca del usuario en diferentes situaciones, por ejemplo, que es de gran interés para aseguradoras como Maphre a la hora de proponerle un seguro de vida más adaptado, y sobre todo con una prima más ajustada a la realidad) o de ocio y restauración (al conocer por dónde se suele correr, días de la semana en que se corre, hora del día,

UNIVERSIDAD DE GRANADA
abiertaugr la formación abierta de la UGR
CC BY NC ND

etc., cabe la posibilidad de ofrecerle servicios de restauración pre o post actividad deportiva, adaptados en contenido y precio al tipo de actividad que se suele realizar).

En este sentido siempre hay que tener en cuenta quién es el propietario de los datos, respetando todos los derechos de los usuarios en lo relativo a la posterior explotación de los mismos por parte de la empresa fabricante del dispositivo conectado, en exclusiva o en colaboración con otras empresas.

5. Conclusiones

En esta unidad se ha querido resaltar la importancia de los datos generados por los servicios conectados para tener un conocimiento mucho más preciso de la experiencia de consumo de los clientes de la empresa. Esto debe servir para una mejor adaptación posterior de los productos y servicios que se ofrecen a los consumidores que contribuyan a un mayor añadido final. En este contexto se ha destacado el papel de los servicios para incrementar ese valor, a lavez que para contribuir a diferenciar el producto/servicio de la empresa. De especial interés desde el punto de vista del consumidor es su nuevo papel como co-creador a través de esa transmisión de información gracias a los dispositivos conectados, lo que además de mejorar lasrelaciones empresa-consumidor, tiene también un papel determinante en la satisfacción final del mismo. Por último, se ha querido destacar la importancia que tienen los datos generados por los diferentes dispositivos conectados para ofrecer mejores productos y servicios al consumidor por otras empresas con o sin base tecnológica, siempre que los consumidores acepten que esos datos se usen con esa finalidad.

6. Bibliografía

[1] I. C. L. Ng y S. Y. L. Wakenshaw, «The Internet-of-Things: Review and research directions», *International Journal of Research in Marketing*, vol. 34, n.º 1, pp. 3-21, mar. 2017,doi: 10.1016/j.ijresmar.2016.11.003.

[2] A. G. Woodside y S. Sood, «Vignettes in the two-step arrival of the internet of thingsand its reshaping of marketing management's service-dominant logic», *Journal of MarketingManagement*, vol. 33, n.º 1-2, pp. 98-110, ene. 2017, doi: 10.1080/0267257X.2016.1246748.

[3] H. Saarijärvi, H. Karjaluoto, y H. Kuusela, «Customer relationship management: the evolving role of customer data», *Marketing Intelligence & Planning*, vol. 31, n.º 6, pp. 584-600,ene. 2013, doi: 10.1108/MIP-05-2012-0055.

[4] Y. Yoo, R. J. Boland, K. Lyytinen, y A. Majchrzak, «Organizing for Innovation in theDigitized World», *Organization Science*, vol. 23, n.º 5, pp. 1398-1408, 2012.

[5] S. L. Vargo y R. F. Lusch, «Evolving to a new dominant logic for Marketing», *Journal of Marketing*, vol. 68, n.º 1, pp. 1-17, 2004.

UNIVERSIDAD DE GRANADA
abiertaugr la formación abierta de la UGR
CC BY NC ND

[6] Y. Yoo, K. J. Lyytinen, R. J. Boland, y N. Berente, «The Next Wave of Digital Innovation: Opportunities and Challenges: A Report on the Research Workshop "Digital Challenges in Innovation Research"», Social Science Research Network, Rochester, NY, SSRN Scholarly PaperID 1622170, jun. 2010. doi: 10.2139/ssrn.1622170.

[7] M. E. Porter y J. E. Heppelmann, «How Smart, Connected Products Are Transforming Competition», *Harvard Business Review*, p. 52, 2014.

[8] «Strava, ¿merece la pena?», *Sportlife*. https://www.sportlife.es/sport- tech/suscripcion-strava-merece-la-pena-segmentos_207511_102.html (accedido ene. 13, 2022).

UNIVERSIDAD DE GRANADA
abiertaugr la formación abierta de la UGR
CC BY NC ND

Módulo 6. IMPACTO DE IOT EN LOS NEGOCIOS

6.2 CONTRIBUCIÓN DE IOT A LA INTELIGENCIA DEL MARKETING. EL MARKETING DE LAS COSAS

Por Juan Sánchez Fernández

Departamento de Comercialización e Investigación de Mercados. Universidad de Granada.

---

1. Introducción

Los datos procedentes de los dispositivos conectados a Internet son de indudable utilidad para apoyar las estrategias de marketing alimentando los sistemas de inteligencia, además sin los costes habituales de obtención de información a través de las clásicas encuestas o las entrevistas cualitativas, ya que en la inmensa mayoría de los casos se transfieren de forma automática, consentida y gratuita por parte de los usuarios de los productos conectados.

La generación de datos que posibilita los objetos conectados abre nuevas formas de afrontar el mercado que pueden llegar a ser sensiblemente diferentes a las tradicionales. De hecho, habrá formas clásicas de operar que quedarán obsoletas y dejarán de ser competitivas a la vez que emergerán nuevas vías de negocio con posibilidades de serlo, fundamentadas en gran parte en nuevas y mejores opciones de servicio ofrecidas de forma exclusiva o, en la mayoría de los casos, compartida por varias empresas. Efectivamente, la mejor manera de afrontar el mercado normalmente no va a ser de forma independiente, sino que la empresa debe tomar consciencia de que pertenece a un nuevo ecosistema donde las posibilidades de negocio son muy diversas e implican a otros muchos agentes, algunos de los cuales pueden ser hasta competidores tradicionales.

Por tanto, los mercados en el contexto de IoT se pueden entender en una lógica de complementariedad o competencia, no siendo tan determinantes las fronteras tradicionales de la industria, sino que lo verdaderamente importante son los espacios del consumidor a los que sirve. En esos espacios se entiende mucho mejor cómo empresas muy diferentes pueden trabajar conjuntamente. Por ejemplo, dentro del bienestar que se busca en cualquier hogar, hay diferentes componentes que contribuyen al mismo, como agua, luz, seguridad, higiene, entretenimiento, etc., donde se abren espacios en los que diferentes empresas pueden crear conjunta o complementariamente valor para el consumidor. Y es que, aunque las necesidades humanas realmente no han cambiado, sí que lo ha hecho la forma, el tiempo, el lugar y la manera en las que son satisfechas. En este contexto, la innovación que ahorre tiempo y mejore eficiencia y la experiencia, terminará imponiéndose [1].

Por tanto, es muy recomendable que la empresa no piense solo en términos de integración vertical u horizontal, sino en innovación tangencial para servir a sus propios consumidores a través de

UNIVERSIDAD DE GRANADA
abiertaugr la formación abierta de la UGR
CC BY NC ND

diferentes ofertas en el mismo espacio de consumo, muchas de las cuales pueden ser fruto de estas alianzas y coordinaciones, usando para ello el vínculo de los datos que pueden generar los objetos conectados comercializados por empresas de muy distinta índole.

En los apartados que siguen abordaremos el tratamiento de la competencia desde una perspectiva amplia que nos permitirá comprender mejor los aspectos determinantes de la competitividad empresarial en un contexto de IoT. Así mismo, se recomendará una metodología para la concreción de un modelo de negocio basado en IoT con posibilidades de ser rentable y sostenible. Por último, se reflexiona sobre los diferentes roles que lo agentes que operan en este nuevo contexto poder desempeñar, que evidentemente no solo pasan por ser fabricante o consumidor final.

2. Internet de las cosas y competitividad

La naturaleza cambiante de los productos en la línea de lo comentado en la sección anterior, así cómo las nuevas cadenas de valor, impulsan a las empresas a reestructurar prácticamente todo lo que venían haciendo a nivel interno. Efectivamente, los productos conectados pueden alterar la estructura de las industrias y la propia naturaleza de la competencia, enfrentando a las empresas a nuevas amenazas, pero también a interesantes oportunidades, afectando al diseño, fabricación y comercialización de estos productos [2].

No obstante, aunque IoT cambiará muchas cosas, los planteamientos sobre las reglas de la competencia y la generación de ventaja competitiva seguirán siendo bastante parecidas a las tradicionales. Para comprenderlo mejor, siguiendo a Porter & Heppelmann (2014), nos vamos a apoyar en el concepto de rivalidad ampliada, que establece que la competitividad de una empresa depende de cinco fuerzas fundamentales:

- Poder de negociación de los clientes
- Rivalidad actual
- Amenaza de nuevos entrantes
- Productos sustitutivos
- Poder de negociación de los proveedores

En los próximos epígrafes analizamos en detalle cada una de esas cinco fuerzas, planteando ejemplos que ayuden a comprender mejor el efecto de los productos conectados sobre la posición competitiva de las empresas.

En lo que hace referencia al **poder de negociación** de los clientes, los productos conectados incrementan de manera significativa sus posibilidades de diferenciación, fundamentalmente a través de nuevos servicios que mejoren la experiencia general disfrutada por los usuarios de los mismos y basada en la información procedente de los datos que la propia experiencia de consumo genera. Una mayor diferenciación, que mejora la satisfacción vía una mejor experiencia, de forma paralela conlleva una reducción de la sensibilidad al precio por parte del consumidor. Efectivamente, el conocimiento mucho más preciso de cómo usan los consumidores los productos, posibilita una segmentación mucho más precisa, establecer

UNIVERSIDAD DE GRANADA
abiertaugr la formación abierta de la UGR
CC BY NC ND

políticas de precios menos basados en los costes y que recojan mejor el valor aportado al consumidor por el producto/servicio, así como añadir nuevos servicios/utilidades que generen valor adicional a los productos que se consumen [2]. Por supuesto, todo esto implica una relación empresa-consumidor mucho más estrecha, lo que en última instancia puede suponer que los costes de cambio de proveedor para los clientes se incrementen significativamente, o lo que viene a ser lo mismo, se reduzca el poder de negociación de los clientes.

Garmin, la empresa que venimos empleado desde la sección anterior como ejemplo, no para de incluir nuevas prestaciones en algunos de sus dispositivos, como la función de música, sistema que permite descargar y almacenar contenido de audio en el mismo reloj, escuchando música cuando se quiera sin necesidad de que el smartphone esté dentro de alcance, por lo que móvil ya no será una molestia para salir a correr. No obstante, si queremos una base musical mucho más rica, siempre se puede instalar la funcionalidad de Spotify a través del servicio Garmin IQ, también comentado en la sección anterior, que ofrece la posibilidad de escuchar listas de reproducción sin necesidad de conexión. Incluso sugerir listas de reproducción en función del tiempo y el ritmo de la actividad desempeñada por el cliente.

Figura 1: *Garmin con Spotify*

Por el contrario, los productos conectados también pueden incrementar el poder de negociación de los clientes, al disponer estos de una mejor comprensión del rendimiento del producto, reduciendo la dependencia del fabricante para consejos/servicios, pudiendo enfrentar a un productor con otro buscando nuevos servicios o reducciones en el precio. Además, en este camino hacia la lógica dominante de servicio frente a modelos basados en la propiedad física, ya introducida en la sección anterior, puede llegar a ser mucho más fácil para con consumidor plantearse un cambio de proveedor ya que los costes necesarios se pueden reducir significativamente, o lo que viene a ser lo mismo, se incrementa su poder de negociación.

A pesar de la gran cantidad de servicios adicionales que ofrece Garmin a través de su aplicación Garmin Connect, y de la gran cantidad de datos que los usuarios han podido ir registrando de sus actividades físico-deportivas en sus servidores de datos, que en principio podría llegar a suponer una barrera de salida importante para dichos usuarios, en los últimos años la irrupción de empresas como Strava han contribuido reducir esas barreras de salida, al ser directamente trasladables esos datos de Garmin a Strava, con lo que cambiar de dispositivos GPS y servicios de información

UNIVERSIDAD DE GRANADA
abiertaugr la formación abierta de la UGR
CC BY NC ND

asociados es relativamente más fácil ahora, pues la inmensa mayoría de los principales marcas de GPS son también compatibles con Strava.

Para tratar de solucionar este tipo de problemáticas, el desarrollo de funcionalidades propias y exclusivas por parte de la empresa puede ayudar a reducir esa facilidad de cambio de proveedor y, por tanto, ese poder de negociación de los usuarios.

En relación a esto, una de las primeras cuestiones que se debe hacer la empresa en este sentido es si debería seguir un sistema abierto o cerrado en relación a la programación empleada en el software de los productos conectados. Optar por un sistema abierto permite integrar/añadir diferentes servicios/funcionalidades de distintos proveedores en un mismo sistema o plataforma para conseguir mayor valor añadido para el consumidor, una mayor satisfacción y, por extensión, unos mayores costes de cambio. No obstante, también es cierto que un sistema cerrado podría suponer una ventaja competitiva más sostenible para una empresa al controlar y optimizar todas las partes de un sistema.

En lo relativo a la rivalidad entre competidores actuales, los productos conectados tienen la posibilidad de cambiar la rivalidad abriendo numerosas formas de diferenciación y servicios de valor añadido. Se abre la vía de adaptar la oferta para atender a nuevos segmentos de mercado, incluso la posibilidad de llegar a la personalización, mejorando aún más la diferenciación y la determinación del precio [2]. El sistema Garmin IQ a través de una serie de aplicaciones propias o diseñadas por desarrolladores externos permite una personalización minuciosa de sus dispositivos. Algunas de esas aplicaciones son gratuitas, pero otras tienen un coste. Esto abre vías de diferenciación e ingresos adicionales que protegen a Garmin frente a competidores directos como Polar o Suunto.

Figura 2: *Principales competidores de Garmin*

La empresa debería plantearse qué conjunto funcionalidades de los productos conectados debe considerar, para lo cual habría que evaluar cuáles son las que verdaderamente añaden valor a sus clientes, así como los costes marginales asociados a las mismas. Por supuesto las características elegidas deberían contribuir a potenciar la diferenciación y el posicionamiento en el mercado.

Además, habría también que decidir si se desarrollan internamente esas capacidades o se cuenta para ello con organizaciones externas. En el contexto de IoT cada vez es más común la subcontratación de funciones, si bien habría que intentar desarrollar internamente aquellas que ofrecen más potencial de innovación y, por tanto, de ventaja competitiva.

Por otra parte, los nuevos entrantes en un contexto de productos conectados se enfrentan a obstáculos significativos, comenzando por los altos costes asociados al diseño más complejo de los productos y la integración de la propia tecnología en los mismos. Las barreras de entrada también aumentar cuando los pioneros obtienen ventajas clave al ser los primeros en recopilar y analizar

UNIVERSIDAD DE GRANADA
abiertaugr la formación abierta de la UGR
BY NC ND

datos sobre los productos con los que posteriormente mejorarlos y rediseñarlos, añadir nuevas prestaciones o servicios, así como mejorar el servicio posventa. Esto puede aumentar la lealtad de los clientes y los costes de cambio, aumentando también las barreras de entrada.

Sin embargo, las barreras de entrada se pueden reducir, con la aparición de competidores que pueden prestar gracias a los datos nuevos servicios a los consumidores sin necesidad de trabajar con productos físicos. La empresa debería plantearse si le conviene entrar en nuevos negocios monetizando los datos de sus productos a través de su venta a otras empresas para que puedan prestar servicios complementarios, aunque también sustitutivos en algunos casos.

Respecto a la amenaza de productos sustitutivos, los productos conectados pueden ofrecer un rendimiento, adaptación y valor superior al consumidor que los productos tradicionales sustitutivos, reduciendo las amenazas de sustitución y mejorando la rentabilidad y crecimiento de la empresa.

No obstante, en muchas industrias, los productos inteligentes crean nuevas amenazas de sustitución como consecuencia de la ampliación de las capacidades de los productos convencionales. Aplicaciones como Strava, ya comentada, u otras como Wikiloc, Nike Run Club o Google Fit compiten con Garmin sin necesidad de desarrollar dispositivos físicos, sino que basan su cartera de servicios en app que se instalan directamente en los smartphones.

Figura 3: *Competidores virtuales de Garmin*

En cuanto al poder de negociación de los proveedores, el software de los productos conectados reduce las necesidades de adaptaciones físicas y, por lo tanto, también el número de variedades de componentes físicos. Por tanto, la importancia de los proveedores de componentes físicos se reduce significativamente, del mismo modo que su poder de negociación. También pueden aparecer nuevos proveedores que anteriormente no eran necesarios: proveedores de sensores, software, conectividad, sistemas operativos integrados, almacenamiento de datos, análisis, y otras partes de carácter tecnológico. Google, Microsoft o Amazon y otros gigantes tecnológicos tienen el talento y las capacidades que muchos fabricantes no tienen y están llegando a ser esenciales en la diferenciación y coste de los productos conectados. El poder de negociación de estos nuevos proveedores puede llegar a ser muy elevado, acaparando una parte importante de la rentabilidad que el fabricantes obtiene de la venta del producto. A modo de ejemplo, algunas de las funcionalidades cartográficas más novedosas que se emplean en Garmin se basan en Google Maps, a la vez que las bases de datos se almacenan en empresas como Amazon (Amazon Web Services- AWS) o Microsoft (Azure).

UNIVERSIDAD DE GRANADA
abiertaugr la formación abierta de la UGR
CC BY NC ND

Figura 4: *Proveedores de Garmin*

No obstante, también es cierto que la empresa se puede plantear si desintermediar total o parcialmente los canales de distribución o las redes de servicio, ya que los productos conectados hacen posible un contacto más intenso y directo con los consumidores que puede conllevar la posibilidad de prescindir de intermediarios, como por ejemplo los encargados del servicio postventa, aumentando en este sentido el poder de negociación de la propia empresa.

3. Internet de las cosas y modelos de negocio

Un modelo de negocio es una descripción del valor que una empresa ofrece a uno o varios segmentos de mercado a través de una organización interna y coordinación externa con otras empresas para crear y comercializar ese valor de forma que genere una corriente de ingresos rentable y sostenible en el tiempo.

Entre otras posibles opciones metodológicas, resulta interesante reflexionar sobre la forma de formular modelos de negocio rentables que creen valor para el consumidor utilizado el CANVAS [3]. Una representación de las 9 cuestiones clave, así como planteamientos sobre los que habría reflexionar/concretar se recoge en la figura 5.

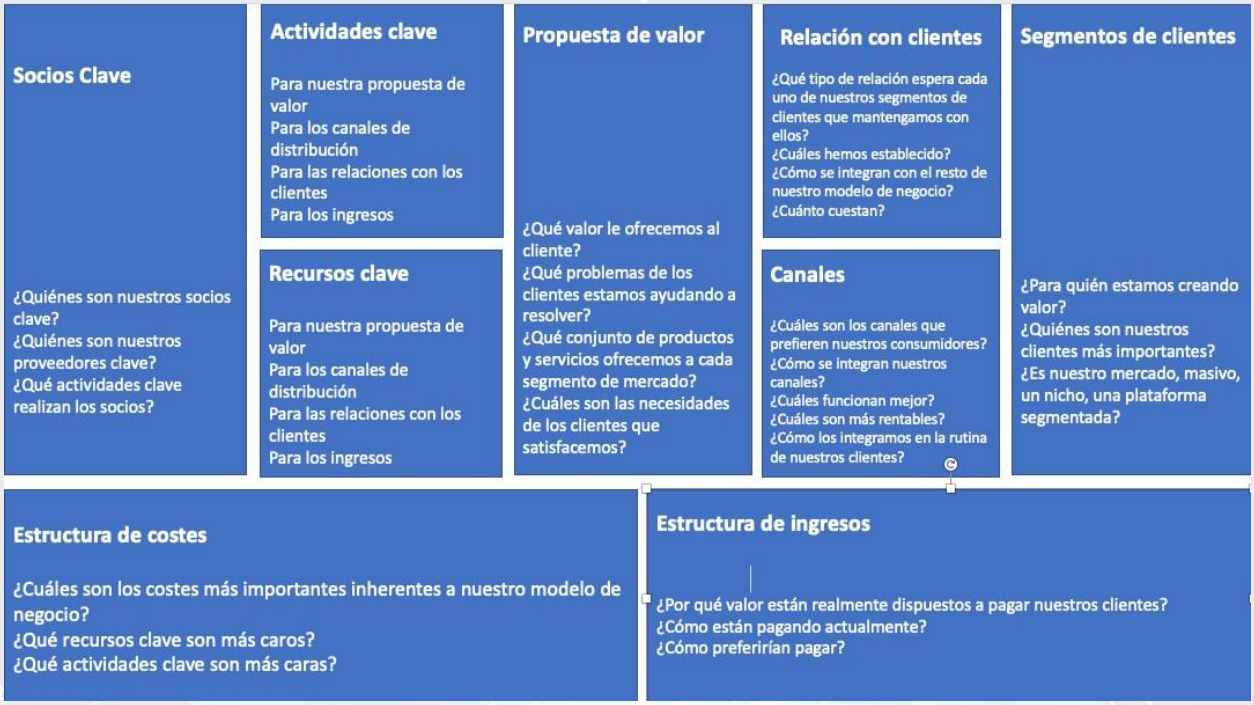

Figura 5: *Modelo Canvas para la compresión de modelos de negocio (basado en [3])*

UNIVERSIDAD DE GRANADA
abiertaugr la formación abierta de la UGR
CC BY NC ND

En el contexto de IoT, no todos los bloques tienen la misma importancia. En un reciente estudio [4], se concluye que los tres que cobran más importancia son: la proposición de valor, la relación con el cliente y la participación de socios clave, lo cual es perfectamente lógico por la especial importancia que tienen los datos en esos tres ámbitos. Más concretamente dentro de la proposición de valor, los aspectos especialmente importantes tienen que ver con la conveniencia, el rendimiento, el desempeño, la comodidad y la posibilidad de actualizaciones. En la relación con los clientes, destaca fundamentalmente la co-creación de valor y la generación de comunidades, así como la posibilidad de ser autosuficientes, al menos en parte. Por otro lado, en el apartado de los socios cobran especial importancia los desarrolladores de app y software, los creadores de hardware y los analistas de datos.

En esta misma línea es interesante la aportación de Clauss, quien desarrolló una conceptualización tridimensional de modelo de negocio con base en IoT que se resumen en la tabla siguiente [5].

Tabla 1: *Innovación de modelos de negocio para soluciones IoT*

Innovación en la creación de valor	Nuevas competencias	Nuevos conocimientos y habilidades técnicas Nuevos conocimientos y habilidades de marketing
	Nuevas tecnologías/equipamiento	Capacidades tecnológicas nuevas Software y equipamiento nuevo
	Nuevos socios	Nuevos socios tecnológicos Nuevos socios especializados en ofrecer servicios complementarios
	Nuevos procesos	Proceso de desarrollo de nuevas soluciones en IoT Proceso de creación de nuevos mercados
Innovación en la proposición de valor	Nuevas ofertas	Servicios de análisis Nuevos servicios a través de modelos compartidos
	Nuevos clientes y mercados	Nuevo enfoque de mercado Demanda de nuevos clientes
	Nuevos canales	Diseño de canales directos nuevos
	Nuevas relaciones con los clientes	Personalización Colaboración e interacción continua Co-creación
Innovación en la captura de valor	Nuevos modelos de ingresos y costes	Modelo de producto como servicio
		Soluciones de bajo coste Soluciones escalables

Fuente: [5]

UNIVERSIDAD DE GRANADA
abiertaugr la formación abierta de la UGR
CC BY NC ND

En este sentido, se abre un nuevo abanico de modelos de negocio para captar valor, desde una visión del modelo de propiedad tradicional en el que el cliente se beneficia de nueva eficiencia en el servicio hasta un modelo de producto-servicio en el que el fabricante conserva la propiedad y asume la responsabilidad de los costes de funcionamiento y servicio del producto a cambio de un pago periódico o por consumo. Del mismo modo, se abre también la posibilidad de los productos compartidos, una utilización más eficiente de los productos que son consumidos intermitentemente (como el ejemplo de las bicicletas ya comentado en la Unidad 1).

4. Dispositivos Roles empresariales en un ecosistema basado en IoT

Por tanto, las posibilidades de negocio en el contexto de IoT son muy diversas para las empresas. Se pueden distinguir cuatro grandes roles en el ecosistema de IoT [6]:

- Habilitadores: aquellas empresas que desarrollan, implementan y mantienen la infraestructura tecnológica fundamental del IoT. Son fundamentalmente empresas tecnológicas.
- Adoptadores: empresas que aplican las tecnologías de IoT para mejorar sus operaciones y optimizar sus procesos empresariales (entre las que se encuentran las empresas de fabricación), pero que no ofrecen una solución de IoT propia para sus clientes.
- Innovadores: empresas que crean sus propios productos/servicios conectados. Diseñan, desarrollan, integran y ofrecen servicios de IoT a los clientes. Por ejemplo, al conectar una reloj deportivo a Internet el fabricante obtiene una gran cantidad de datos sobre el uso del aparato.
- Potenciadores: son las empresas que diseñan sus propios servicios de valor añadido, además de los servicios prestados por los innovadores, que son exclusivos del IoT. Pueden ofrecer servicios que amplíen el valor añadido por los fabricantes de productos.

En un contexto general de co-creación de servicios, es necesario no sólo definir los roles, sino también crear entornos de apoyo para la integración de recursos. Esto puede hacerse "centrándose en (1) mecanismos que faciliten las interacciones entre diversos actores, (2) adaptando los procesos internos para acoger a los diferentes actores (roles), y (3) aumentar la transparencia de las actividades de integración de recursos en el ecosistema de servicios" [7]. Y es que los límites de la competencia se amplían para abarcar un grupo de productos relacionados que, en conjunto, satisfacen una necesidad subyacente más amplia. La función de un producto se optimiza con otros productos relacionados. Por tanto, la base de la competencia cambia desde la funcionalidad de un producto individual al rendimiento de un sistema de producto más amplio, donde la empresa es solo uno de los actores.

Los hogares inteligentes, que incluyen entre otras cosas iluminación, climatización, entretenimiento, sanidad y seguridad son otro buen ejemplo. Los productos serán diseñados para que tengan el mayor

UNIVERSIDAD DE GRANADA
abiertaugr la formación abierta de la UGR
CC BY NC ND

impacto posible en el sistema, proporcionando el máximo valor añadido dentro de un ecosistema, buscando efectos sinérgicos, donde el resultado final agregado es superior a la suma del valor de las diferentes partes por separado.

En este contexto es muy probable que surjan nuevas empresas, sin necesidad de grandes activos, que aprovechen las posibilidades de los productos conectados para crear valor, llegándose en algunos acasos a estrategias "sin producto", donde la principal ventaja estará precisamente en el sistema que conecta los productos, no en los propios productos [2]. El caso de Strava, ya comentado anteriormente, es un claro ejemplo [8].

En cualquier caso, con independencia de cuál sea el rol que se adopte, a nivel de organización la forma de conseguir ventaja competitiva seguirá recayendo en la capacidad de diferenciación que permita un precio superior, menores costes que los competidores, o ambas cosas combinadas. Esto no cambia en el contexto IoT.

5. Conclusiones

En esta unidad, se ha realizado un repaso de como el contexto de IoT puede afectar a los modelos de negocio. Se ha reflexionado de las posibles amenazas y oportunidades que IoT pueden suponer para la competitividad empresarial. Así mismo se ha recomendado una

metodología para la determinación de modelos de negocio con base IoT. Aunque surgen nuevos roles y nuevas formas de relacionarse las empresas gracias a los datos generados por los objetos conectados, las claves de la competitividad siguen siendo fundamentalmente las mismas, diferenciación (en este caso fundamentalmente basadas en los servicios) y coste.

6. Bibliografía

[1] I. C. L. Ng y S. Y. L. Wakenshaw, «The Internet-of-Things: Review and research directions», International Journal of Research in Marketing, vol. 34, n.o 1, pp. 3-21, mar. 2017, doi: 10.1016/j.ijresmar.2016.11.003.

[2] M. E. Porter y J. E. Heppelmann, «How Smart, Connected Products Are Transforming Competition», Harvard Business Review, p. 52, 2014.

[3] A. Osterwalder y Y. Pigneur, Business model generation: a handbook for visionaries, game changers, and challengers, vol. 1. John Wiley & Sons, 2010.

[4] R. M. Dijkman, B. Sprenkels, T. Peeters, y A. Janssen, «Business models for the Internet of Things», International Journal of Information Management, vol. 35, n.o 6, pp. 672-678, dic. 2015, doi: 10.1016/j.ijinfomgt.2015.07.008.

[5] T. Clauss, «Measuring business model innovation: conceptualization, scale development, and proof of performance», R&D Management, vol. 47, n.o 3, pp. 385-403, 2017, doi: 10.1111/radm.12186.

UNIVERSIDAD DE GRANADA
abiertaugr la formación abierta de la UGR
CC BY NC ND

[6] T. Saarikko, U. H. Westergren, y T. Blomquist, «The Internet of Things: Are you ready for what's coming?», Business Horizons, vol. 60, n.o 5, pp. 667-676, sep. 2017, doi: 10.1016/j.bushor.2017.05.010.

[7] R. Lusch y S. Nambisan, «Service Innovation: A Service-Dominant Logic Perspective», MIS Q., 2015, doi: 10.25300/MISQ/2015/39.1.07.

[8] «Strava Similar App Development | Develop A Social Fitness Network App». https://www.trootech.com/strava-similar-app-development/ (accedido ene. 13, 2022).

UNIVERSIDAD DE GRANADA
abiertaugr la formación abierta de la UGR
CC BY NC ND

Módulo 6. IMPACTO DE IOT EN LOS NEGOCIOS

6.3 LA IMPORTANCIA DEL DATOS Y LA TRANSFORMACIÓN DE LOS DATOS PROCEDENTES DE IOT EN INTELIGENCIA DE MARKING.

Por Salvador del Barrio García

Departamento de Comercialización e Investigación de Mercados. Universidad de Granada.

---

1. Introducción

En esta unidad vamos a abordar el estudio del dato como fuente de inteligencia para los negocios y el marketing. Comprenderemos el significado que tienen los datos en la actualidad, y como estos por sí mismos carecen de valor si no se organizan, limpian, tratan y analizan para convertirlos en información valiosa que genere conocimiento para la toma de decisiones para las empresas.

Será importante también entender las diferentes formas en las que se organizan los datos, según estos sean estructurados, no estructurados o semiestructurados, y las posibilidades que permiten en cada caso.

También se analizará en esta unidad desde una perspectiva práctica la importancia actual que tienen los datos en los negocios y en el marketing, describiendo múltiples ejemplos de cómo la innovación en los datos ha generado beneficios para la asistencia médica, el transporte, el medioambiente, la economía, la agricultura y el marketing.

Finalmente, se analizará el proceso de transformación de los datos en inteligencia, describiendo las distintas fases que es necesario implementar, recurriendo a diversos ejemplos.

2. El dato como fuente de inteligencia de negocios y marketing

En un mundo globalizado y tan digitalizado como el actual, los datos se han convertido en el nuevo "petróleo" del siglo XXI [1], o tal y como apuntan algunos expertos en un futuro próximo llegará a asemejarse más bien al "agua", al ser esencial para la vida de los ciudadanos, de las empresas y de la sociedad, en su conjunto. Pero al igual que el agua, los datos deben estar accesibles y limpios para que sean realmente útiles para la toma de decisiones.

En módulos anteriores se ha explicado la diferencia entre dato, información y conocimiento. El dato se refiere a símbolos o conjuntos de símbolos utilizados para representar o describir un valor numérico, un objeto o una idea. Los datos son imprescindibles para el diseño de estrategias y para la toma de decisiones en las empresas, pero por sí solos carecen de valor. Es necesario procesarlos, organizarlos y analizarlos para convertirlos en información útil.

UNIVERSIDAD DE GRANADA
abiertaugr la formación abierta de la UGR
CC BY NC ND

A modo de ejemplo, y para que se entienda mejor, en gastronomía los datos serían los ingredientes y la información sería los platos ya cocinados [2]. Con los mismos ingredientes se pueden conseguir diferentes platos en función a diferentes elaboraciones, técnicas y procedimientos. Con los datos pasa lo mismo, la aplicación de diferentes herramientas, técnicas y procedimientos a los datos van a dar lugar a informaciones diversas para generar inteligencia (conocimiento) de negocio y de marketing. Ese conocimiento o inteligencia permitirá detectar información clave y relevante (*insights*) para la toma de decisiones.

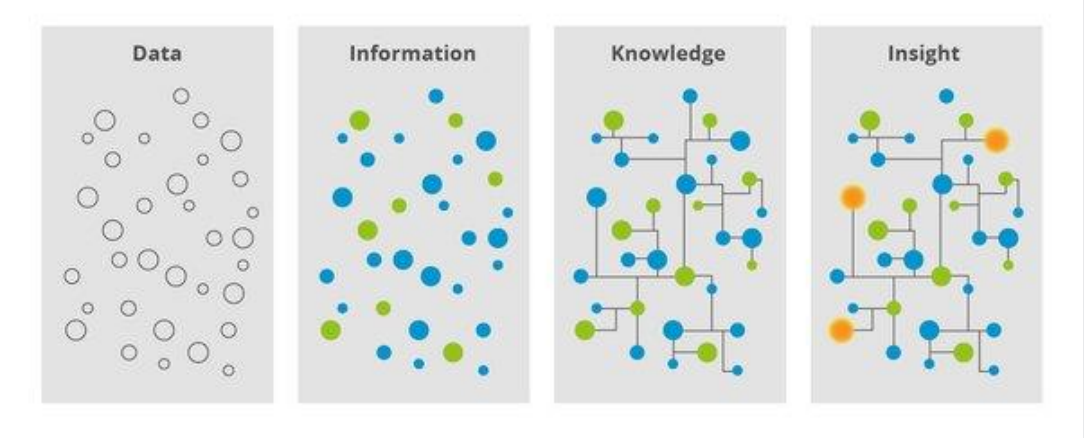

Figura 3.1. *Evolución del dato hacia la inteligencia y la toma de decisiones [3]*

Hoy en día, cada segundo que pasa se generan y recopilan millones de datos procedentes de clientes cuando interactúan con los productos y servicios de las empresas (ventas, llamadas a servicios de atención al cliente, petición de información…), resultados de estudios de consumidor, lectura de escáneres (terminales TPV en comercios, audiencias en medios), datos de accesos y clics en páginas web y anuncios, acciones que realizan los usuarios con sus *smartphones* (descarga de Apps…), comentarios (*reviews*) a productos, o datos de sensores IoT, entre otros.

Llegados a este punto es importante conocer que existen diferentes tipos de datos que se dividen según su estructura en: (1) estructurados, (2) no estructurados y (3) semiestructurados.

Datos estructurados

Los datos estructurados están organizados y dispuestos en filas y columnas, donde las filas representan los casos o sucesos y las columnas las variables, y son el tipo de datos con el que la mayoría de empresas trabaja. Son los que podemos encontrar en una base de datos relacional o una simple hoja de clientes, donde las filas representan los distintos clientes y las columnas sus características: dirección, teléfono, email, fecha de alta como cliente, número de productos adquiridos en el último mes, promedio del valor de las ventas realizadas, etc. Este tipo de bases de datos relacionales permiten ingresar, buscar y manipular los datos con relativa rapidez y se gestionan mediante un tipo de lenguaje de programación estructurado conocido como SQL.

UNIVERSIDAD DE GRANADA
abiertaugr la formación abierta de la UGR
CC BY NC ND

Los datos estructurados pueden ser generados por las personas, como por ejemplo los datos de registro de clientes, de comportamiento de compra o de flujo de clics en Internet y redes sociales; o por las máquinas, como los datos procedentes de sensores IoT, de registro web, generados en el punto de venta mediante escáneres o datos financieros.

ID	IMPLICA	CONCIM	AMA	Exp_PAG_3	sexo_m	edad_m	estado_c	resid	frecuenc	frecuencia 2 niveles	frecuencia tres niveles	Comenzado visita	Tratamiento	Autenticidad
895	2,30	2,500	2,429	Alta	Hombre	Entre 25 y 40 años	Soltero	Granada Provincia	2-5 veces	Alta	Media	Si	T4	Baja
277	2,70	3,000	3,000	Alta	Hombre	Entre 25 y 40 años	Soltero	Granada Capital	2-5 veces	Alta	Media	Si	T6	Alta
387	4,50	4,250	4,429	Alta	Mujer	Menor de 25 años	Soltero	Granada Provincia	2-5 veces	Alta	Media	Si	T4	Baja
597	3,40	4,250	3,571	Media	Hombre	Entre 25 y 40 años	Soltero	Granada Capital	2-5 veces	Alta	Media	Si	T4	Baja
528	4,30	3,000	3,286	Alta	Mujer	Menor de 25 años	Soltero	Granada Capital	2-5 veces	Alta	Media	Si	T4	Baja
424	2,90	5,000	2,571	Baja	Mujer	Entre 41 y 55 años	Casado	Resto de Andalucía	Alta	Alta	Alta	Si	T4	Baja
229	1,20	2,500	3,000	Media	Mujer	Entre 25 y 40 años	Soltero	Resto de España	2-5 veces	Alta	Media	Si	T4	Baja
225	1,10	3,000	2,143	Baja	Hombre	Entre 25 y 40 años	Soltero	Resto de España	2-5 veces	Alta	Media	Si	T6	Alta
266	2,10	1,750	4,000	Alta	Hombre	Más de 55 años	Casado	Granada Provincia	2-5 veces	Alta	Media	Si	T5	Media
235	2,90	3,500	4,429	Alta	Mujer	Entre 41 y 55 años	Casado	Granada Capital	1 vez	Baja	Baja	Si	T6	Alta
224	2,60	3,250	3,286	Baja	Mujer	Entre 25 y 40 años	Soltero	Extranjero	2-5 veces	Alta	Media	No	T5	Media
216	3,10	4,000	2,857	Baja	Mujer	Entre 25 y 40 años	Casado	Resto de España	1 vez	Baja	Baja	Si	T6	Alta
246	4,70	3,500	3,571	Media	Mujer	Entre 25 y 40 años	Soltero	Resto de Andalucía	Alta	Alta	Alta	Si	T5	Media
210	2,60	3,250	1,571	Alta	Mujer	Entre 25 y 40 años	Soltero	Resto de Andalucía	Alta	Alta	Alta	Si	T5	Media
223	3,10	3,750	3,143	Alta	Mujer	Entre 25 y 40 años	Soltero	Resto de Andalucía	2-5 veces	Alta	Media	Si	T4	Baja
247	2,90	2,750	3,143	Baja	Mujer	Más de 55 años	Casado	Resto de España	1 vez	Baja	Baja	No	T6	Alta
228	3,40	3,000	3,714	Media	Mujer	Menor de 25 años	Soltero	Resto de España	1 vez	Baja	Baja	Si	T4	Baja
193	2,70	3,250	1,286	Alta	Hombre	Entre 25 y 40 años	Soltero	Granada Provincia	1 vez	Baja	Baja	No	T5	Media
307	4,20	3,000	3,143	Alta	Mujer	Entre 25 y 40 años	Soltero	Resto de Andalucía	2-5 veces	Alta	Media	Si	T5	Media
230	1,80	2,250	3,714	Media	Mujer	Menor de 25 años	Soltero	Granada Capital	1 vez	Baja	Baja	Si	T5	Media
576	3,60	3,000	3,571	Alta	Hombre	Entre 25 y 40 años	Soltero	Resto de Andalucía	2-5 veces	Alta	Media	Si	T6	Alta
503	4,40	4,750	3,571	Media	Mujer	Entre 25 y 40 años	Soltero	Resto de España	2-5 veces	Alta	Media	Si	T4	Baja
7984	5,00	4,500	4,714	Baja	Mujer	Entre 41 y 55 años	Otras situaciones	Resto de España	Alta	Alta	Alta	Si	T5	Media
331	4,50	3,500	4,429	Alta	Hombre	Entre 41 y 55 años	Casado	Resto de Andalucía	2-5 veces	Alta	Media	Si	T6	Alta
755	3,80	3,500	3,571	Media	Hombre	Entre 25 y 40 años	Soltero	Granada Provincia	1 vez	Baja	Baja	Si	T5	Media
580	3,50	3,250	3,429	Alta	Mujer	Entre 25 y 40 años	Soltero	Granada Capital	1 vez	Baja	Baja	Si	T6	Alta
291	1,60	3,000	2,714	Alta	Mujer	Entre 41 y 55 años	Casado	Resto de Andalucía	2-5 veces	Alta	Media	Si	T5	Media
249	2,10	2,250	2,857	Media	Hombre	Entre 41 y 55 años	Casado	Resto de España	2-5 veces	Alta	Media	Si	T4	Baja
253	3,10	3,500	2,714	Media	Mujer	Menor de 25 años	Soltero	Resto de España	2-5 veces	Alta	Media	Si	T4	Baja
425	1,00	4,000	4,143	Baja	Mujer	Entre 25 y 40 años	Soltero	Resto de España	2-5 veces	Alta	Media	Si	T5	Media
196	1,40	1,750	1,857	Media	Mujer	Menor de 25 años	Soltero	Resto de España	2-5 veces	Alta	Media	Si	T5	Media
189	3,70	3,250	3,429	Baja	Hombre	Menor de 25 años	Soltero	Resto de España	1 vez	Baja	Baja	Si	T6	Alta
184	3,60	4,000	2,286	Baja	Mujer	Entre 25 y 40 años	Soltero	Resto de España	2-5 veces	Alta	Media	Si	T5	Media
192	1,00	1,750	3,000	Baja	Hombre	Entre 25 y 40 años	Otras situaciones	Resto de España	2-5 veces	Alta	Media	Si	T5	Media
186	4,70	3,500	4,143	Media	Mujer	Entre 25 y 40 años	Soltero	Resto de España	2-5 veces	Alta	Media	Si	T6	Alta
213	1,50	2,250	1,714	Baja	Hombre	Entre 25 y 40 años	Otras situaciones	Resto de España	1 vez	Baja	Baja	Si	T4	Baja
362	2,10	2,500	3,714	Media	Mujer	Menor de 25 años	Soltero	Resto de España	1 vez	Baja	Baja	Si	T6	Alta
269	4,50	5,000	5,000	Baja	Mujer	Entre 25 y 40 años	Casado	Resto de España	2-5 veces	Alta	Media	Si	T4	Baja
201	3,10	3,000	3,571	Media	Mujer	Entre 25 y 40 años	Soltero	Resto de España	más de 10 veces	Alta	Alta	Si	T4	Baja
1987	3,00	3,500	3,143	Alta	Mujer	Entre 25 y 40 años	Soltero	Granada Capital	2-5 veces	Alta	Media	Si	T5	Media
710	2,50	3,250	3,857	Baja	Hombre	Entre 25 y 40 años	Soltero	Granada Capital	2-5 veces	Alta	Media	Si	T5	Media
216	4,30	3,000	3,429	Alta	Hombre	Entre 25 y 40 años	Soltero	Extranjero	Alta	Alta	Alta	Si	T4	Baja
302	3,00	2,500	3,714	Media	Mujer	Entre 25 y 40 años	Soltero	Resto de España	2-5 veces	Alta	Media	Si	T6	Alta
151	1,10	2,750	3,714	Baja	Hombre	Entre 25 y 40 años	Soltero	Granada Capital	1 vez	Baja	Baja	Si	T5	Media

Figura 2. *Ejemplo de datos estructurados procedentes de una encuesta a clientes. En las filas el número de identificación (ID) del cliente y en las columnas las distintas variables recogidas para cada cliente.*

Datos no estructurados

Los datos no estructurados no tienen una estructura predefinida lo que hace más complicado su análisis y visualización mediante herramientas convencionales. Por tanto, no se organizan en bases de datos relacionales y requieren de otro tipo de herramientas para su administración y análisis. Están formados en su mayoría por textos, imágenes o vídeos que se organizan en bases de datos no relacionales que no usan tablas sino documentos que recogen todos los atributos. Estas bases son mucho más flexibles a la hora de crear esquemas de información, permiten una mayor escalabilidad y volúmenes de datos y son muy funcionales permitiendo mediante APIs exclusivas poder acceder a ellas para extraer la información que se requiera en cada momento.

En la actualidad entre el 80% de todos los datos de las empresas son no estructurados [4] y dicho porcentaje seguirá incrementándose con el tiempo a medida que se vaya extendiendo la digitalización a todos los niveles en la empresa y se vaya extendiendo el IoT.

Analizar los datos no estructurados es mucho más complejo que los estructurados, dado que requiere sistemas mucho más potentes que permitan manejar un mayor volumen de datos y herramientas de análisis más sofisticadas que trabajen con datos masivos de carácter no solo cuantitativo sino también cualitativo, como textos o imágenes. En este sentido, las técnicas de

UNIVERSIDAD DE GRANADA
abiertaugr la formación abierta de la UGR
BY NC ND

minería de datos y de inteligencia artificial aplicadas a este tipo de datos constituyen hoy en día una importante ventaja competitiva para las empresas al permitir, por ejemplo, entender y predecir el comportamiento real de los consumidores observando sus hábitos, o sus sentimientos hacia un producto o servicio examinando sus opiniones y revisiones que hacen.

Entre las fuentes que generan más datos no estructurados hoy en día cabe destacar las redes sociales que se han convertido en parte de nuestras vidas y a las que acceden miles de millones de personas en todo el planeta. Estas redes generan ingentes cantidades de datos segundo a segundo las 24 horas en forma de texto, imágenes, videos, audios, ubicaciones geográficas, etc. Asimismo, el IoT es una fuente enorme de generación de datos mediante sensores instalados en los automóviles, en productos electrónicos o dispositivos accesorios o wearables, entre otros.

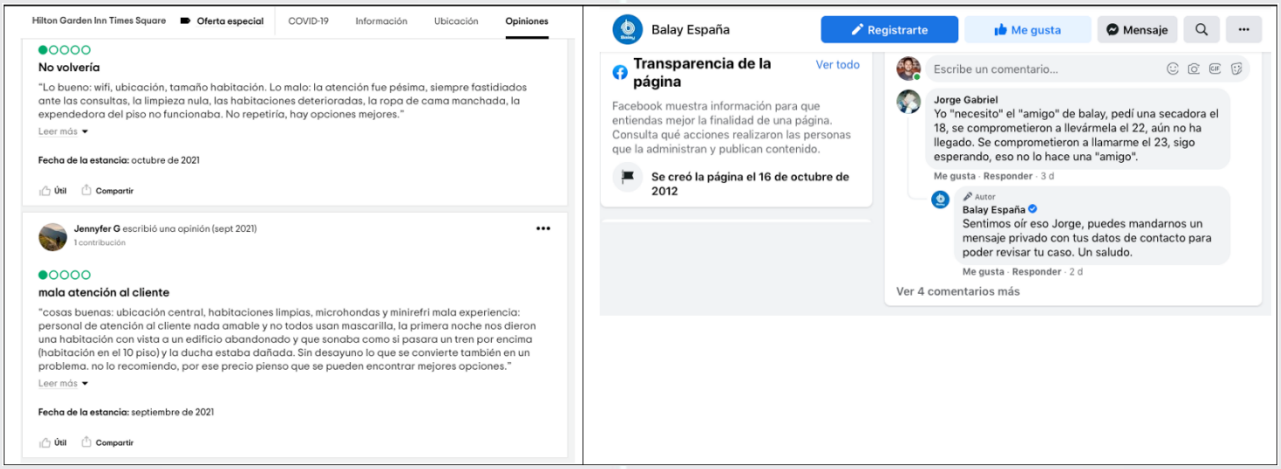

Figura 3. *Ejemplo de datos no estructurados procedentes de comentarios de usuarios de un hotel en el portal TripAdvisor o de opiniones en redes sociales de clientes.*

Datos semiestructurados

Los datos semiestructurados están a medio camino entre las dos tipologías anteriores, al no estar completamente estructurados como los que se registran en bases relacionales, si bien proporcionan otros medios para describir dichos datos y las jerarquías entre ellos, como por ejemplo las etiquetas semánticas. El típico ejemplo de datos semiestructurados sería el código HTML que impone jerarquía de los datos a través de los elementos semánticos, aunque no restringe la cantidad de información a incluir en un documento. Otros ejemplos serían los datos de los correos electrónicos y archivos adjuntos, o los archivos comprimidos [5].

La forma que adoptan este tipo de datos es en árbol con etiquetas. Contienen también metadatos (etiquetas y elementos) que se emplean para agruparlos y describir cómo se almacenan, aunque su gestión y almacenamiento es algo más difícil que los estructurados.

UNIVERSIDAD DE GRANADA
abiertaugr la formación abierta de la UGR
CC BY NC ND

Figura 4. *Ejemplo de datos semiestructurados procedentes del código fuente HTML de una página web*

3. Infraestructura La importancia actual del dato en los negocios y en el marketing

Como se ha comentado anteriormente, la principal fuente de generación de datos en la actualidad es el IoT. El IoT está revolucionando el mundo que nos rodea. Un reciente informe del *Internacional Data Corporation* (IDC) [6] estima que para 2025 habrá más de 40 mil millones de dispositivos basados en IoT que generarán alrededor de 175 zettabytes de datos y más de 6000 millones de personas interactuando con esos datos.

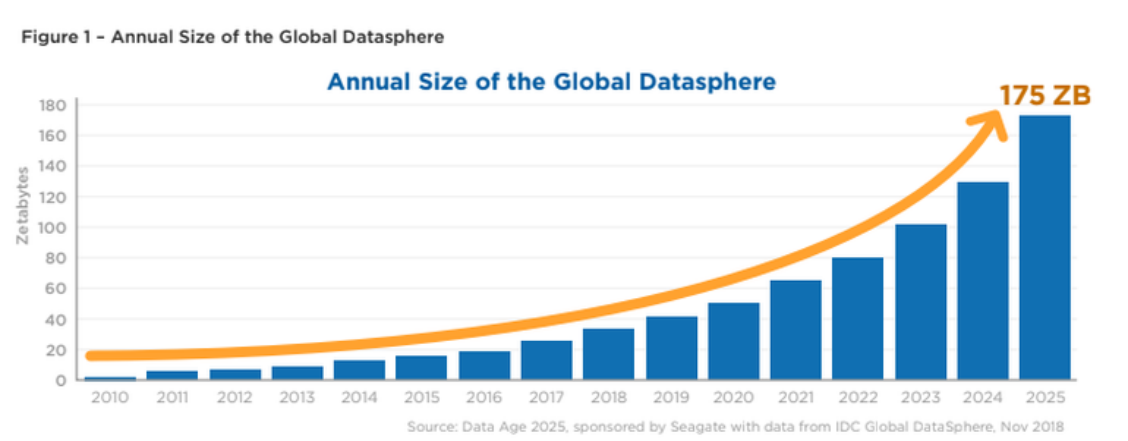

Figura 5. *Evolución del crecimiento del volumen de datos en el mundo*

IoT se está volviendo cada vez más esencial para la innovación, la adaptación y el éxito en las empresas, especialmente para aquellas con grandes capacidades de conectividad y acceso a la Red [8]. Las empresas están encontrando gradualmente nuevas formas de adoptar y aprovechar los datos que genera el IoT en varias funciones de marketing, particularmente en todas las actividades orientadas a las relaciones con el cliente. Además, a medida que los datos empiezan a abundar y su coste de almacenamiento baja, las empresas están apostando por la implementación de

UNIVERSIDAD DE GRANADA
abiertaugr la formación abierta de la UGR
CC BY NC ND

herramientas de ciencia de datos que dejen al descubierto conocimiento valioso a partir de enormes cantidades de datos. Por tanto, nos dirigimos hacia un mundo de información casi ilimitada donde las organizaciones y empresas usarán los datos que generamos para realizar predicciones que mejoren la vida de los ciudadanos y de los consumidores. Por ejemplo, los datos de tráfico permitirán sincronizar los semáforos de una manera más eficaz y gestionar de manera más eficiente el transporte público y privado para hacer ciudades más inteligentes, o los dispositivos accesorios (wearables) tales como los relojes inteligentes nos ayudarán a realizar un seguimiento de nuestro estado de salud y permitirán mejorar nuestra actividad física y bienestar.

Como señala la Fundación BSA (software.org), las oportunidades que los datos, y la innovación de los datos a través de las herramientas de software disponible, ofrecen en la actualidad a las empresas es enorme. Los datos se están convirtiendo en nuevos productos en sí, soluciones e innovaciones que nos están cambiando la vida, además de que están generando una nueva economía que podría aportar al PIB global para el 2030 15 billones de $ [9]. Los efectos económicos de los datos se están extendiendo a muchos sectores económicos y, sin lugar a dudas, está proporcionando importantes incrementos en la productividad de las empresas. Así, diversos estudios han puesto de manifiesto que las empresas que están orientando su toma de decisiones guiadas por los datos están incrementando sus niveles de productividad en torno al 6%.

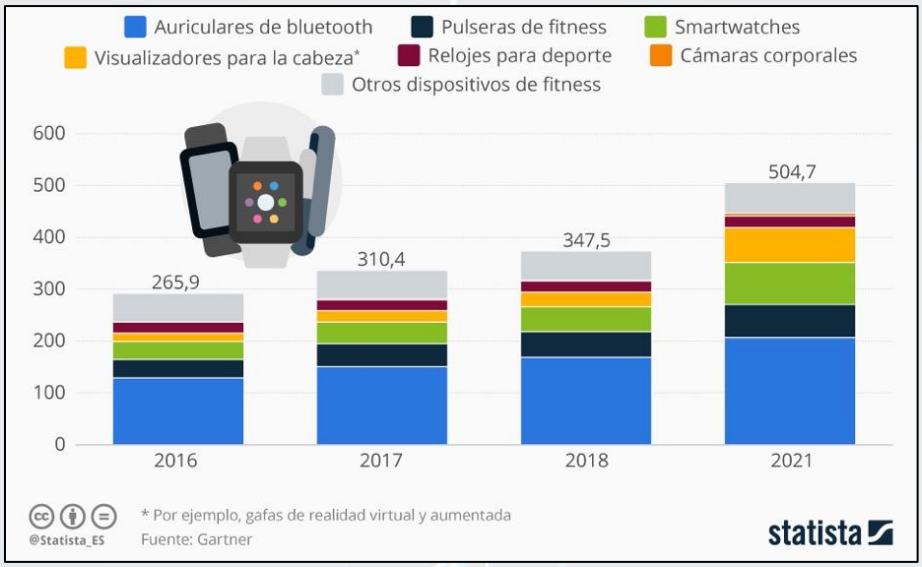

Figura 6. *El futuro del mercado de los dispositivos wearables (en millones de unidades) [10]*

Además, la innovación de los datos no debe ser vista únicamente en términos de su contribución a la productividad y valor de las empresas, sino también en términos de creación de empleo para la sociedad. El mundo en torno al dato ya está creando cientos de miles de empleos en todo el mundo directa e indirectamente en forma de analistas de datos, desarrolladores de software, gestores de sistemas, ejecutivos de datos, empresas de hardware y de dispositivos y sensores, etc. Asimismo, los beneficios para la sociedad en general y los ciudadanos son incuestionables:

UNIVERSIDAD DE GRANADA
abiertaugr la formación abierta de la UGR
BY NC ND

- Asistencia médica: la innovación de los datos ayuda a las personas a vivir más tiempo y tener vidas más saludables.
- Transporte: la gestión innovadora de datos permite ahorrar tiempo y dinero en las distintas formas de transporte a la vez que ahorra combustible y salva vidas.
- Medioambiente: la gestión masiva y eficiente de datos está permitiendo la reducción del consumo de energías fósiles y el incremento de energías limpias, lo que contribuye al cuidado del medioambiente.
- Sector financiero: la innovación de los datos aumenta la eficiencia en la gestión de los recursos financieros en empresas y entidades financieras, reduce el fraude y mejora la rentabilidad.
- Agricultura: la innovación de los datos mejora la productividad de las explotaciones agrícolas y produce mejores alimentos con menos uso de recursos.

Figura 7. *Algunos ejemplos de beneficios para la sociedad del uso e innovación de los datos [9]*

UNIVERSIDAD DE GRANADA
abiertaugr la formación abierta de la UGR
CC BY NC ND

Los datos que generan el IoT ofrecen infinitas oportunidades a los especialistas en marketing. Permite a las empresas brindar soluciones de productos y servicios que hasta ahora no se habían explotado o eran difíciles, así como escuchar y responder a las necesidades de los clientes, con el mensaje correcto, en el momento adecuado, y a través de los canales adecuados. Algunos ejemplos de estas nuevas posibilidades que ofrecen los dispositivos IoT al marketing serían:

- Altavoces inteligentes, asistentes personales y hogares inteligentes. Estos dispositivos no solo pueden actuar como un asistente personal sino que también pueden administrar y controlar otros dispositivos IoT conectados. Algunos estudios han demostrado que este tipo de dispositivos pueden ahorrar hasta un 10% en calefacción y un 15% en refrigeración [11].
- IoT en automóviles. El automóvil es uno de los productos de consumo más prometedores en lo que respecta a las capacidades de IoT. Desde automóviles semiautónomos gestionados con IoT, hasta automóviles que avisan sobre averías o mantenimientos.
- Servicio al cliente basado en IoT. Aunque es incipiente en comparación con otras aplicaciones de IoT, el uso de IoT en el servicio al cliente es cada vez más utilizado. Desde *chatbots* inteligentes en sitios web y aplicaciones móviles hasta pulseras inteligentes (*magic bands*) que permiten acceder a múltiples servicios. La tecnología permite resolver muchas de las quejas de los clientes de forma automática y brinda a las empresas la oportunidad de estar disponibles 24 horas al día, 7 días a la semana.
- Dispositivos médicos inteligentes y atención médica. La industria médica y de la salud ha sido uno de los primeros beneficiarios de IoT. Específicamente, las capacidades de monitoreo remoto de IoT han sido de gran utilidad para pacientes, cuidadores, médicos y hospitales. La monitorización remota a través de estos dispositivos también ha llevado a la reducción de la duración de la estancia hospitalaria, lo que se traduce en ahorros de costes.

Figura 8. *Servicios que ofrece la pulsera inteligente de Disney [12]*

UNIVERSIDAD DE GRANADA
abiertaugr la formación abierta de la UGR
CC BY NC ND

4. Dispositivos La transformación del dato en inteligencia de negocios y marketing

Como se ha comentado anteriormente, los datos por sí mismo carecen de valor si no son sometidos al tratamiento, análisis e interpretación de los mismos. Así pues, hoy en día se hace más necesario que nunca transformar los datos de que se dispone en respuestas, es decir, en conocimiento e inteligencia que permita a las organizaciones dar respuesta a las cuestiones que se le plantean y tomar las mejores decisiones.

En este proceso de transformación de los datos intervienen 4 grandes fases secuenciales: recopilación, almacenamiento, análisis e interpretación de los datos.

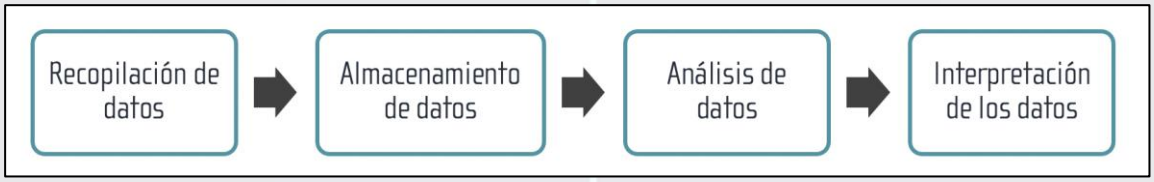

Figura 9. *Fases del proceso de transformación de los datos en inteligencia (elaboración propia)*

Fase I. Recopilación de datos

A lo largo de la historia las empresas han ido recopilando datos para facilitar la gestión de sus negocios y ayudarles a la toma de decisiones. Sin embargo, los datos que existían eran escasos, más allá de los que controlaba la propia empresa a nivel interno, y las posibilidades para recolectar datos del entorno o de los consumidores era muy costosa en recursos y tiempo. Hoy en día las empresas tienen la suerte de poder extraer grandes cantidades de datos del mundo que les rodea como consecuencia del desarrollo de la tecnología. Por tanto, los datos han dejado de ser un recurso escaso para convertirse en un recurso fundamental, renovable y cada vez más abundante.

Algunos ejemplos que muestran la situación actual en torno al mundo del dato serían los siguientes [13]:

- Los usuarios de YouTube suben más de 48 horas de video cada minuto.
- Cada minuto se crean más de 570 páginas web en el mundo.
- Las marcas y organizaciones reciben en Facebook más de 35000 likes cada minuto.
- Google registra más de 3,5 millones de búsquedas diariamente.
- El volumen de datos empresariales en el mundo se dobla cada 1,2 años aproximadamente.

Fase II. Almacenamiento de datos

Otro aspecto clave en la revolución en torno al dato que estamos viviendo en los últimos años es el abaratamiento del coste de almacenamiento de los datos. Si en los años 80 almacenar un gigabyte de datos costaba cientos de miles de dólares, en la actualidad ese coste se ha reducido por más de un

UNIVERSIDAD DE GRANADA
abiertaugr la formación abierta de la UGR
CC BY NC ND

factor de 10 millones hasta costar no más de 1 céntimo. Para ponerlo en contexto, es como si hoy en día pudiéramos conducir un automóvil casi 10.000 veces alrededor de la tierra con lo que se pagaba en los 80 por 3,7 litros de gasolina [9].

Un aspecto que sin duda ha contribuido al abaratamiento de los costes de almacenamiento ha sido el desarrollo de la tecnología en la nube que ha abierto un mundo de posibilidades para las empresas y los ciudadanos, al poder de manera fácil, rápida y flexible poder almacenar enormes cantidades de datos. Esta circunstancia ha hecho además que ya no sea necesario eliminar datos antiguos para poder grabar nuevos, sino que los datos antiguos se unen a los nuevos para poder crear modelos predictivos mucho más precisos y eficientes mediante herramientas de inteligencia artificial.

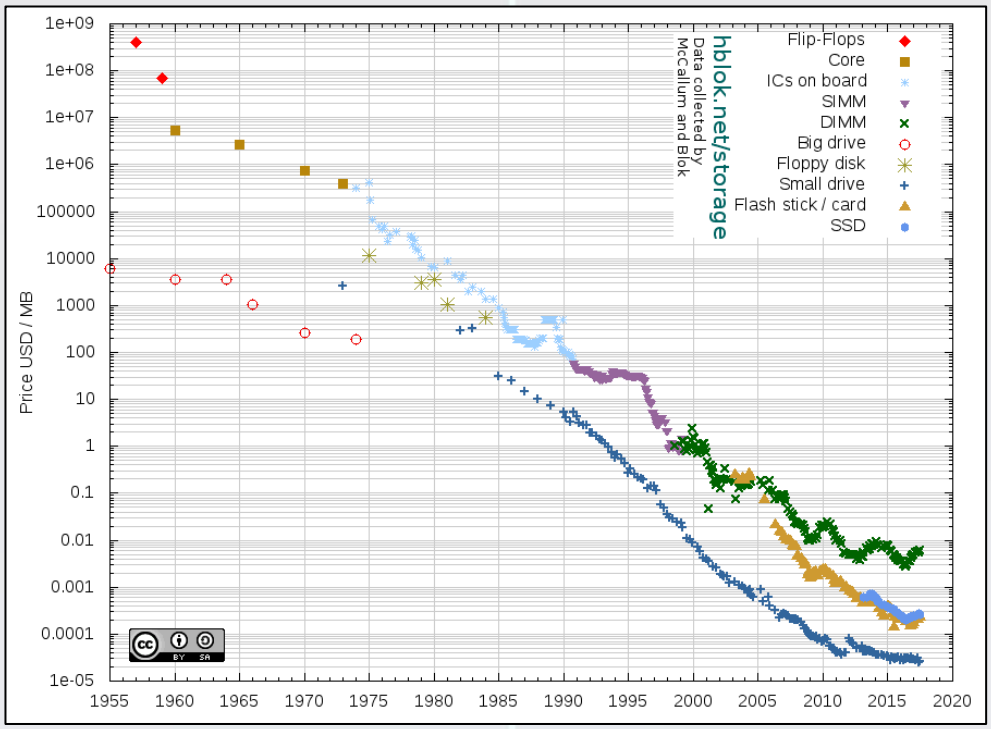

Figura 10. *Evolución del precio de almacenamiento de un megabyte según dispositivos [14]*

Fase III. Análisis de datos

El crecimiento de los datos y la reducción de los costes de almacenamiento se ha visto acompañado por el desarrollo tecnológico en términos de hardware y software que está transformando significativamente la disciplina del análisis de datos. Hace varias décadas los analistas de datos disponían de herramientas estadísticas y softwares clásicos para enfrentarse a bases de datos estructuradas limitadas en cuanto al volumen de datos. Era difícil que un analista de datos manejara volúmenes de miles de registros.

Hoy en día, los científicos de datos (como ahora se les denomina) se enfrentan a todo tipo de datos (estructurados, no estructurados y semiestructurados) y a volúmenes de cientos de miles de datos

UNIVERSIDAD DE GRANADA
abiertaugr la formación abierta de la UGR
BY NC ND

muy diversos y procedentes de muy diversas fuentes que precisan de gran cantidad de tiempo para su preprocesamiento y limpieza. Se estima que el tiempo que dedican en la actualidad los científicos de datos para preparar los datos antes del análisis está entre el 50% y el 80% de todo su tiempo [15].

Las herramientas de ciencia de datos están en continuo desarrollo permitiendo a los científicos examinar montañas de datos para encontrar pepitas de información de oro. Así la Minería de Datos y el Big Data, junto con las herramientas de Inteligencia Artificial como el Machine Learning (aprendizaje de máquinas) o el Deep Learning (aprendizaje profundo) y los algoritmos que surgen de las mismas están brindando a los científicos herramientas de vanguardia para comprender cantidades ingentes de datos y poder descifrar el conocimiento tan valiosos que los mismos esconden.

A todo ello se le une el desarrollo de la computación en la nube (*cloud computing*) que ofrecen grandes compañías tecnológicas como Google, Amazon, Microsoft, Alibaba o IBM, y que permiten a los usuarios ejecutar herramientas sofisticadas de análisis con datos almacenados en distintos lugares al mismo tiempo y con un poder de procesamiento enorme.

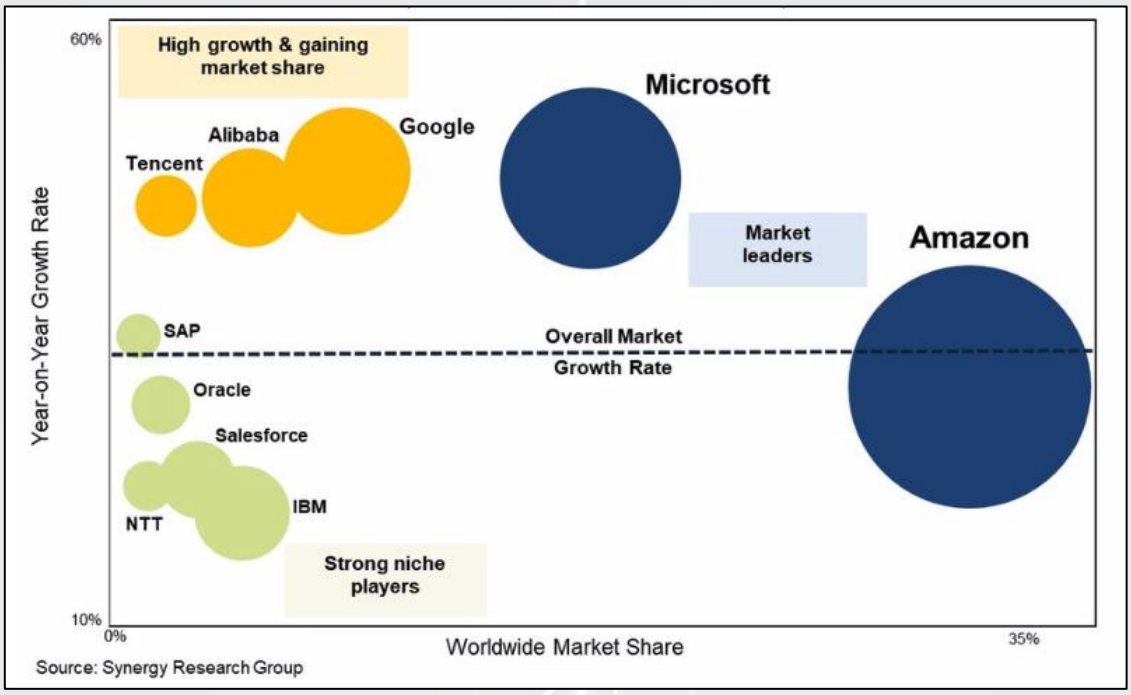

Figura 11. *Cuota de mercado mundial de las empresas proveedoras de computación en la nube (2020) [16]*

Todo esto ayuda a empresas y organizaciones a realizar predicciones cada vez más fiables que les van a ayudar a conocer cómo se comportará el mercado ante determinadas circunstancias y decisiones clave. Un ejemplo de la utilidad del potencial del análisis de datos en tiempo real es la industria del automóvil que están diseñando vehículos cargados de miles de sensores y análisis en tiempo real de la información que generan, lo que permite poder evitar choques, corregir trayectorias, evitar atascos de tráfico o incluso conducir de manera autónoma. Una de las compañías de automóviles

UNIVERSIDAD DE GRANADA
abiertaugr la formación abierta de la UGR
CC BY NC ND

que más está trabajando en esta tecnología es Tesla quien recientemente ha implementado el sistema *Full Self-Driving*, un sistema de conducción autónoma basado en múltiples sensores y cámaras y un potente ordenador que toma decisiones complejas en tiempo real en función a los datos que les proporcionan los sensores [17].

Fase IV. Interpretación de datos

La última fase de todo este proceso de transformación de los datos en inteligencia es la interpretación de los mismos. El desarrollo que se ha producido en los últimos años en cuanto a herramientas y algoritmos de análisis de datos y de procesamiento en la nube debe ir acompañado de un buen conocimiento de toda la información que generan esas herramientas, de cara a realizar una interpretación adecuada que sea realmente útil para la toma de decisiones. Por ello hoy en día a los científicos de datos se les exige una formación interdisciplinar no solo en informática y matemáticas, sino también en otros campos como la ingeniería, la economía y el marketing, la medicina, o la lingüística, para poder encontrar y entender esas pepitas de información de oro que proporcionan todos estos sistemas avanzados aplicados a diversos ámbitos.

Por ejemplo, en los últimos años el avance en los algoritmos de procesamiento de imágenes está teniendo un impacto tremendo en áreas como la medicina al poder avanzar en el diagnóstico precoz de tumores cerebrales o de mama. Durante la crisis de la Covid-19 se han desarrollado algoritmos para poder diagnosticar la enfermedad con una simple ecografía de los pulmones sin tener que recurrir a los test PCR, con los beneficios que ello produce en términos de poder actuar mucho antes en los tratamientos.

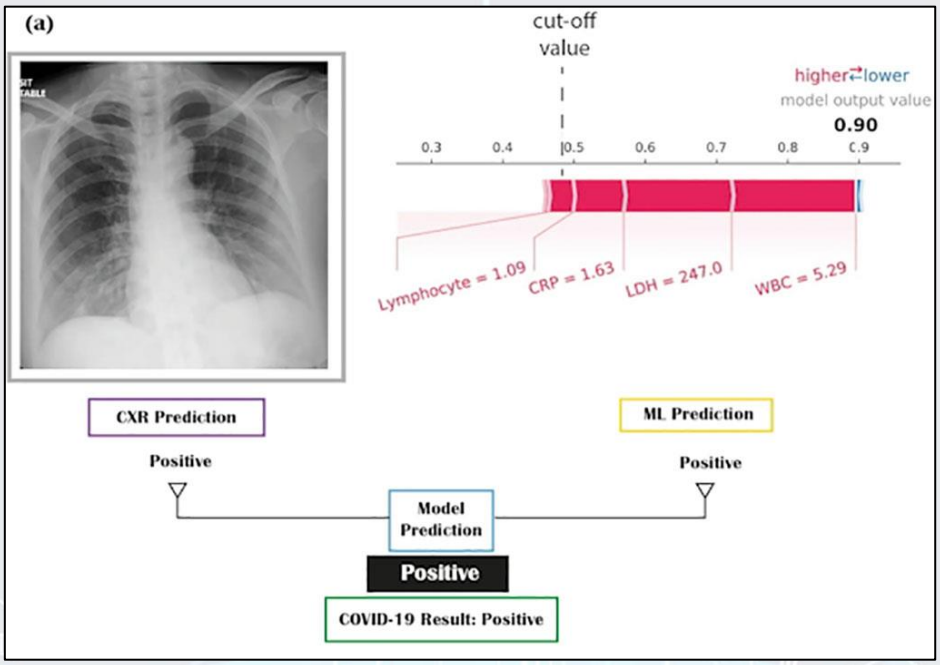

Figura 12. *Modelo de inteligencia artificial que identifica a pacientes Covid-19 mediante análisis de radiografías de tórax [18]*

UNIVERSIDAD DE GRANADA
abiertaugr la formación abierta de la UGR
CC BY NC ND

5. Conclusiones

Tras la lectura de esta unidad el estudiante habrá sido capaz de entender la diferencia entre los datos, la información que surge de ellos y el conocimiento que puede extraerse de los mismos para tomar decisiones. Asimismo, creemos que el lector será más consciente de la importancia que hoy en día tienen los datos para la sociedad en su conjunto, y para las empresas y el marketing, en particular.

El desarrollo de la tecnología de almacenamiento, tratamiento y análisis de datos está abriendo enormes posibilidades a las empresas para el desarrollo de productos y servicios que faciliten la vida a los consumidores y usuarios. En este unidad hemos recogido múltiples ejemplos de las posibilidades que los datos pueden proporcionar para la mejora de la sociedad actual.

En la siguiente unidad se abordará desde una perspectiva muy práctica algunos ejemplos de análisis de datos masivos para que el lector pueda visualizar los beneficios y posibilidades que generan los datos y su tratamiento y análisis adecuado en la actualidad.

6. Bibliografía

[1] https://www.forbes.com.mx/red-forbes-los-datos-son-el-nuevo-petroleo/

[2] https://www.bbva.com/es/cual-es-la-diferencia-entre-datos-e-informacion-como-convertir-el-big-data-en-valor-real-para-las-personas/

[3] https://qph.fs.quoracdn.net/main-qimg-2194a84ea0ec8808f7b03f409f16a989-pjlq

[4] https://www.itop.es/blog/item/big-data-cuales-son-los-datos-no-estructurados-generados-por-maquinas-y-por-las-personas.html

[5] https://blogs.imf-formacion.com/blog/tecnologia/tipos-de-datos-datos-estructurados-semiestructurados-y-no-estructurados-202006/

[6] https://iotbusinessnews.com/2020/08/10/08984-connected-devices-will-generate-79-zettabytes-of-data-by-2025/

[7] https://www.muycomputerpro.com/2018/12/03/datos-en-tiempo-real

[8] X. Yu, B. Nguyen, Y. Chen (2016). Internet of things capability and alliance: Entrepreneurial orientation, market orientation and product and process innovation. *Internet Research*, 26, 402–434.

[9] https://data.bsa.org/wp-content/uploads/2015/10/BSADataStudy_es.pdf

[10] https://es.statista.com/grafico/10842/los-auriculares-de-bluetooth-conquistan-el-mercado/

[11] https://nest.com/thermostats/real-savings/

[12] https://www.disneyadictos.com/2016/01/que-son-las-magicband.html

UNIVERSIDAD DE GRANADA
abiertaugr la formación abierta de la UGR
CC BY NC ND

[13] https://waterfordtechnologies.com/big-data-interesting-facts/

[14] https://hblok.net/storage_data/storage_memory_prices_2005-2017-12.png

[15] https://www.nytimes.com/2014/08/18/technology/for-big-data-scientists-hurdle-to-insights-is-janitor-work.html?_r=0

[16] https://openwebinars.net/blog/tipos-de-cloud-computing/

[17] https://www.xataka.com/vehiculos/tesla-esta-activando-su-conduccion-autonoma-full-self-driving-a-quien-desee-probarla-solo-para-buenos-conductores

[18] https://www.labmedica.es/covid-19/articles/294789062/modelo-de-inteligencia-artificial-ia-identifica-a-los-pacientes-con-covid-19-mediante-pruebas-de-sangre-y-radiografias-de-torax.html

UNIVERSIDAD DE GRANADA
abiertaugr la formación abierta de la UGR
CC BY NC ND

Módulo 6. IMPACTO DE IOT EN LOS NEGOCIOS

6.4 APLICACIÓN DE HERRAMIENTAS DE CIENCIA DE DATOS PARA GENERAR INTELIGENCIA DE MARKETING

Por Salvador del Barrio García

Departamento de Comercialización e Investigación de Mercados. Universidad de Granada.

1. Introducción

Esta unidad temática aborda desde una perspectiva práctica la aplicación de la ciencia de datos para generar inteligencia de negocios y marketing. Tras hacer un repaso de lo que es, las principales herramientas que la componen y su conexión con otros conceptos como la Inteligencia Artificial, Machine Learning y Deep Learning, se describen algunos ejemplos teóricos de aplicación al ámbito del comercio electrónico, la fabricación de productos, la banca, el transporte, las finanzas o la publicidad.

Finalmente, se examinan brevemente las principales herramientas de ciencia de datos que está a disposición de los científicos y se profundiza en tres ejemplos prácticos de aplicación de la minería de textos al análisis de revisiones de productos, del clustering a una base sobre estilos de vida de los clientes y al análisis de datos masivos longitudinales mediante la herramienta DFA.

2. La ciencia de datos y la inteligencia artificial

La gran cantidad de datos que se generan en la actualidad en el ámbito de las empresas procedentes de los dispositivos IoT exigen llevar a cabo un proceso de recopilación, almacenamiento y análisis para poder encontrar el conocimiento oculto en los mismos. Para ello, la ciencia de datos se considera hoy en día como una disciplina fundamental. Uno de los grandes objetivos que persigue es la búsqueda de modelos que describan patrones y comportamientos que ayuden a la toma de decisiones y al diseño de productos y servicios innovadores para satisfacer las necesidades y deseos de los consumidores.

Antes de analizar algunas de las herramientas de ciencias de datos más utilizadas en el ámbito empresarial y de marketing y realizar algún ejemplo práctico, es importante delimitar conceptualmente el término ciencia de datos y diferenciarlo de otros términos relacionados. La ciencia de datos es una disciplina que busca hacer que los datos sean útiles y generen conocimiento al objeto de conseguir ventajas competitivas para las empresas que la practican. Muy vinculado a este término está la inteligencia artificial (IA) que podría definirse como la simulación de procesos de inteligencia humana por parte de máquinas, y que incluye el aprendizaje mediante la adquisición de información y reglas, el razonamiento que usa esas reglas para alcanzar conclusiones y la autocorrección [1].

UNIVERSIDAD DE GRANADA
abiertaugr la formación abierta de la UGR
CC BY NC ND

Las diferencias principales entre ambos términos serían [2]:

- La ciencia de datos es un proceso integral que implica el preprocesamiento, análisis, visualización y predicción; mientras que la IA es la implementación de algoritmos para pronosticar eventos futuros.
- La ciencia de datos se compone de varias técnicas estadísticas, mientras que la inteligencia artificial hace uso de algoritmos informáticos.
- Las herramientas involucradas en la ciencia de datos son mucho más que las que se usan en la inteligencia artificial. Esto se debe a que la ciencia de datos implica varios pasos para analizar datos y generar conocimientos a partir de ellos.
- La ciencia de datos trata de encontrar patrones ocultos en los datos. La IA trata de impartir autonomía al modelo de datos.
- La ciencia de datos busca crear modelos que utilizan conocimientos estadísticos. La IA, en cambio, sirve para construir modelos que emulen la cognición y la comprensión humana.
- La ciencia de datos implica un menor grado de procesamiento científico en comparación con la IA.

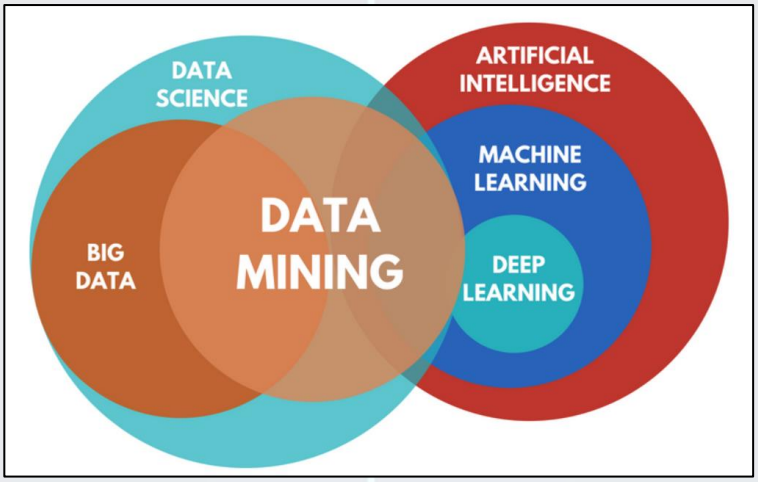

Figura 4.1. *Ciencia de datos vs. IA [3]*

Sin embargo, la parte común de ambas disciplinas está en que giran en torno a los datos y a la necesidad de sacarle el máximo valor, existiendo una fuerte relación o intersección entre ellos. La ciencia de datos es necesaria para la IA ya que esta última necesita grandes cantidades de datos para que aprendan las máquinas, aplicando y validando sus algoritmos y generando inteligencia. Cuantos más datos se disponga y más limpios estén, más podrán aprender las máquinas y más precisos y eficientes serán los resultados. Esto es precisamente lo que conecta la IA con la ciencia de datos, la provisión de "combustible" a la IA, una herramienta poderosa para el científico de datos.

Para entenderlo mejor, la siguiente figura representa la jerarquía de necesidades de la ciencia de datos donde los distintos componentes de la pirámide representarían las operaciones de datos que realizan hoy en día los científicos de datos.

UNIVERSIDAD DE GRANADA
abiertaugr la formación abierta de la UGR
CC BY NC ND

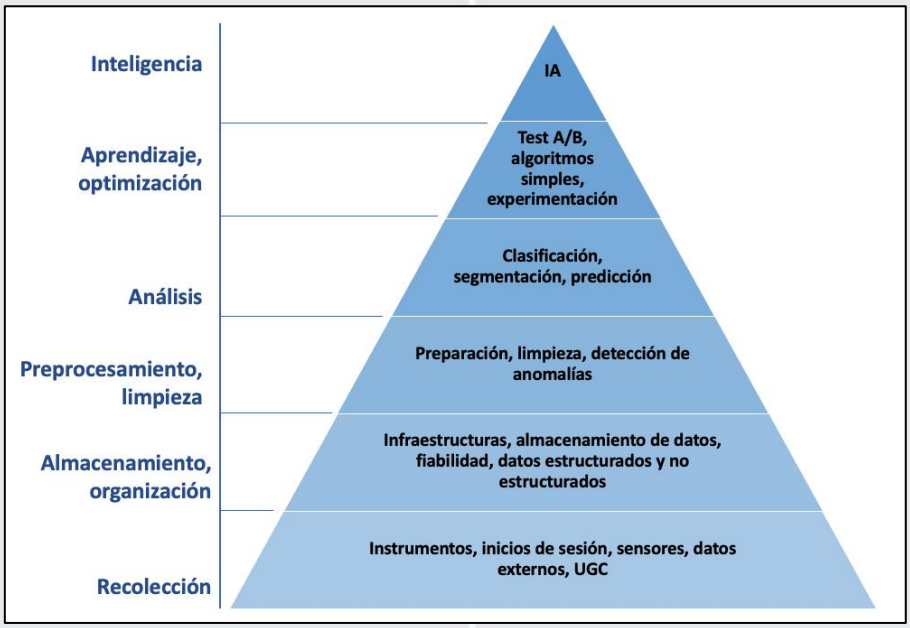

Figura 2. *Jerarquía de necesidades de la ciencia de datos. Fuente: elaboración propia.*

Tanto la ciencia de datos como la IA cuentan con distintas áreas para llevar a cabo sus objetivos. Así, dentro de la ciencia de datos podemos distinguir el Big Data y la Minería de datos (Data Mining) y dentro de la IA podemos diferenciar el aprendizaje de máquinas (Machine Learning) y el Aprendizaje Profundo (Deep Learning).

- El Big Data hace referencia al almacenamiento y procesamiento de grandes volúmenes de datos almacenados en bases de datos, razón por la cual no se relaciona con IA. Esta área busca extraer y almacenar datos de manera sistémica que por lo general son almacenados en "Data Warehouse" que incluyen diversas bases de datos.
- La minería de datos se refiere al análisis de datos para extraer valor de los mismos con resultados exploratorios, descriptivos o predictivos. En este sentido sí que se relaciona con la IA, y más concreto con el Machine Learning, ya que el objetivo es extraer valor de esa "mina" de datos, aunque eso puede llevarse a cabo mediante algoritmos de Machine Learning o no. A modo de ejemplo, simplemente obtener el saldo medio de millones de clientes de una base de datos de un banco ya supondría extraer valor a esos datos masivos sin necesidad de IA. En cambio, si se consigue hacer una predicción del saldo que van a tener el próximo año los clientes mediante el desarrollo de un algoritmo de Machine Learning se estará extrayendo muchísimo más valor dentro de la minería de datos [3].
- El Machine Learning comprende una serie de algoritmos de aprendizaje automático basados en modelos estadísticos y matemáticos tales como árboles de decisión o regresión logística que permiten realizar labores de agrupación, predicción o clasificación.

UNIVERSIDAD DE GRANADA
abiertaugr la formación abierta de la UGR
CC BY NC ND

- El Deep Learning forma parte del Machine Learning, pero va un poco más allá al aplicar algoritmos que funcionan de forma similar a cómo lo hace el cerebro humano mediante la aplicación de sistemas de redes neuronales. Estos algoritmos de última generación se están utilizando en la actualidad para el procesamiento del lenguaje natural (que las máquinas comprendan el lenguaje humano), el procesamiento de imágenes y videos (que las máquinas comprendan el contenido que aparece en las imágenes o videos), sistemas expertos (que las máquinas proporcionen asesoramiento experto) o para comunicar y conectar los distintos dispositivos conectados a través de Internet (IoT).

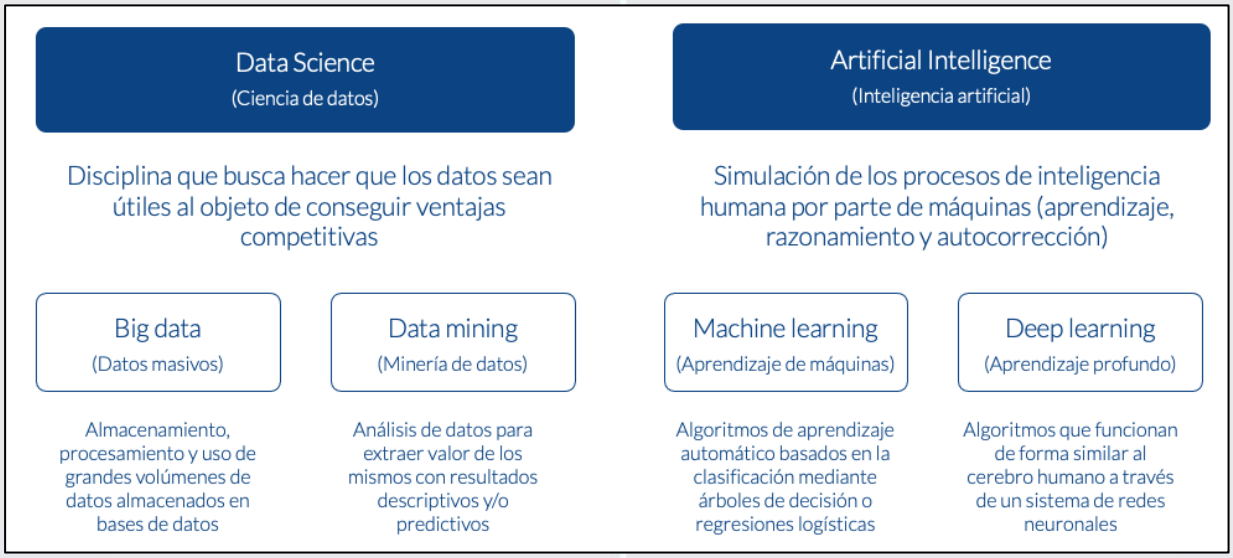

Figura 3. *Áreas que componen la ciencia de datos y la IA. Fuente: elaboración propia.*

3. Aplicaciones de herramientas de ciencias de datos al ámbito del marketing

A continuación, analizaremos desde una perspectiva eminentemente práctica algunas aplicaciones de las herramientas de datos más comunes al ámbito de los negocios y al marketing, no con el objetivo de que el lector aprenda a utilizarlas paso a paso, lo que pertenecería más a un curso sobre ciencia de datos, sino de que se haga una idea del potencial de estas herramientas para sacarle valor a los datos procedentes del mundo de la empresa.

Ejemplos de aplicaciones de ciencia de datos en marketing

Son muchos los ejemplos que se podrían mencionar de aplicaciones de la ciencia de datos al ámbito de los negocios y el marketing:

UNIVERSIDAD DE GRANADA
abiertaugr la formación abierta de la UGR
CC BY NC ND

- Comercio electrónico: para identificar clientes, recomendar productos o analizar *reviews* de productos.
- Fabricación de productos: para predecir problemas potenciales, monitorizar sistemas, diseñar líneas de producción automática, detectar anomalías o programar mantenimiento de las maquinarias.
- Banca: para detectar fraude, establecer valoración de créditos a clientes, predecir la adquisición de productos financieros por parte de los clientes o estimar el valor de un cliente.
- Transporte: para fabricar coches autónomos e inteligentes, mejorar la experiencia del usuario, sistemas de monitorización del vehículo, o la mejora de la seguridad de los pasajeros.
- Finanzas: para segmentar clientes, ayudar en la toma de decisiones estratégicas, análisis de riesgos.
- Publicidad: para la compra de la publicidad programática en Internet y redes sociales, para segmentar a las audiencias, para clasificar a los usuarios, para dirigir la publicidad según los gustos y comportamientos de los consumidores.

Más concretamente, como se muestra en la figura siguiente grandes compañías como Facebook, Amazon, Uber, Bank of America, AirBnb o Spotify han venido usando las herramientas de ciencia de datos de una manera muy exitosa con diferentes propósitos.

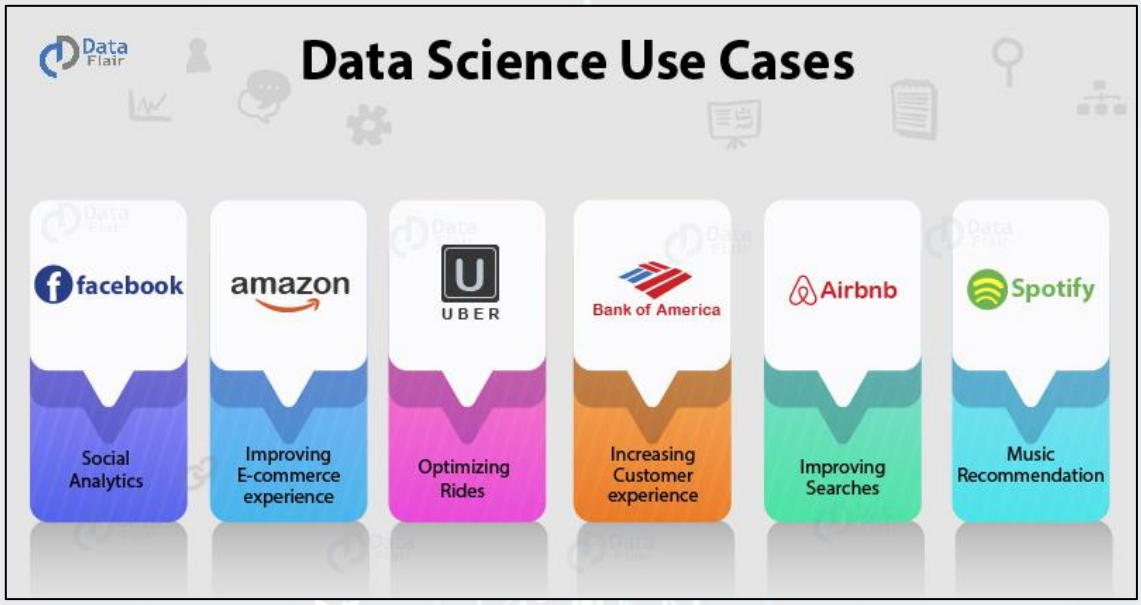

Figura 4. *Ejemplos específicos de compañías multinacionales aplicando ciencia de datos para mejorar la eficiencia de sus negocios [4].*

Amazon ha conseguido su éxito mundial en gran medida por ser capaz de crear una plataforma centrada en el cliente. Se basa en el análisis predictivo para aumentar la satisfacción de sus clientes mediante un sistema de recomendaciones personalizado que se centra en el análisis de las compras

UNIVERSIDAD DE GRANADA
abiertaugr la formación abierta de la UGR
CC BY NC ND

históricas del usuario y de las sugerencias que se extraen de otros usuarios. Asimismo, Amazon tiene un modelo de envío anticipado que utiliza big data para predecir los productos que es más probable que compren sus usuarios, analizando el patrón de sus compras y enviando los productos a su almacén más cercano que puede utilizar en el futuro. Otra área a la que se dirigen todas las plataformas de comercio electrónico es la detección de fraudes. Amazon tiene sus propias formas y algoritmos novedosos para detectar vendedores fraudulentos y compras fraudulentas.

Herramientas de ciencia de datos

Existen una amplia variedad de herramientas que utilizan los científicos de datos para llevar a cabo sus objetivos, entre las más utilizadas podemos destacar las siguientes:

- Clasificación. Son herramientas que permiten predecir variables objetivo del tipo binarias (1/0) o multicategóricas. Entre estas herramientas las más utilizadas son los árboles de decisión, la regresión logística, las redes neuronales, o los algoritmos Naive Bayes, Random Forest o SVM.
- Clustering o agrupamiento. Agrupa o segmenta casos a través del uso de procedimientos de grupos jerárquicos usando reglas, probabilidades y estadísticos. Dentro de este grupo destacan el análisis clúster jerárquico y K-Means y el análisis de clases latentes.
- Regresión. Este tipo de herramientas predicen variables objetivo de tipo numérico. Las más comúnmente utilizadas son los modelos lineales generalizados, la regresión por pasos, las redes neuronales o SVM.
- Reducción de datos mediante extracción de dimensiones. Este tipo de técnicas sirven para reducir el conjunto de variables de una base de datos a un conjunto más reducido de dimensiones que mantienen buena parte de la información contenida en la base original pero más fácil de manejar. En este caso tenemos herramientas como el análisis de componentes principales o el análisis de minería de textos.
- Datos dinámicos (series temporales). Son herramientas que buscan predecir datos numéricos secuenciales a partir de series de datos temporales y resumir tendencias en los mismos. Dentro de estas cabe destacar el análisis de series temporales o el análisis factorial dinámico (DFA).

A continuación, vamos a aplicar de manera práctica y a modo de ejemplo tres de estas herramientas a casos específicos con datos reales típicos de situaciones de marketing: (1) reducción de datos mediante minería de textos para extraer conocimiento de una base de comentarios (*reviews*) sobre vinos realizadas por clientes; (2) clustering para agrupar a un conjunto de clientes de una empresa de los que se dispone de información sobre sus aficiones y comportamientos y (3) análisis de datos dinámicos aplicados a datos de búsqueda de clientes en Google sobre marcas de automóviles.

UNIVERSIDAD DE GRANADA
abiertaugr la formación abierta de la UGR
CC BY NC ND

Aplicación de la herramienta de minería de textos (text mining)

Descripción de la base de datos

Contamos con una base de datos inicial de 129.971 comentarios (*reviews*) realizadas por 19 catadores profesionales de vinos de la guía Wine Enthusiast entre 1999 y 2017. Esta clasificación es publicada en la revista Wine Enthusiast Magazine, una de las revistas más prestigiosas del sector. Esta base incluye además del comentario que hace el experto al vino, otra serie de variables como el país de procedencia del vino, la región de procedencia, la variedad de la uva, el nombre de la bodega y la puntuación cuantitativa que el experto otorga al vino (de 80 a 100 puntos).

	country	description	points	province	taster_name	title	variety	winery
0	Italy	Aromas include tropical fruit, broom, brimstone and dried herb. The palate isn't overly expressive, offering unripened apple, citrus and dried sage alongside br	87	Sicily & Sardinia	Kerin O'Keefe	Nicosia 2013 Vulkà Bianco (Etna)	White Blend	Nicosia
1	Portugal	This is ripe and fruity, a wine that is smooth while still structured. Firm tannins are filled out with juicy red berry fruits and freshened with acidity. It's already drin	87	Douro	Roger Voss	Quinta dos Avidagos 2011 Avidagos Red (Dour	Portuguese Red	Quinta dos Avidagos
2	US	Tart and snappy, the flavors of lime flesh and rind dominate. Some green pineapple pokes through, with crisp acidity underscoring the flavors. The wine was all	87	Oregon	Paul Gregutt	Rainstorm 2013 Pinot Gris (Willamette Valley)	Pinot Gris	Rainstorm
3	US	Pineapple rind, lemon pith and orange blossom start off the aromas. The palate is a bit more opulent, with notes of honey-drizzled guava and mango giving way	87	Michigan	Alexander Peartree	St. Julian 2013 Reserve Late Harvest Riesling (I	Riesling	St. Julian
4	US	Much like the regular bottling from 2012, this comes across as rather rough and tannic, with rustic, earthy, herbal characteristics. Nonetheless, if you think of it a	87	Oregon	Paul Gregutt	Sweet Cheeks 2012 Vintner's Reserve Wild Ch	Pinot Noir	Sweet Cheeks
5	Spain	Blackberry and raspberry aromas show a typical Navarran whiff of green herbs and, in this case, horseradish. In the mouth, this is fairly full bodied, with tomat	87	Northern Spain	Michael Schachner	Tandem 2011 Ars In Vitro Tempranillo-Merlot (N	Tempranillo-Merlot	Tandem
6	Italy	Here's a bright, informal red that opens with aromas of candied berry, white pepper and savory herb that carry over to the palate. It's balanced with fresh acid	87	Sicily & Sardinia	Kerin O'Keefe	Terre di Giurfo 2013 Belsito Frappato (Vittoria)	Frappato	Terre di Giurfo
7	France	This dry and restrained wine offers spice in profusion. Balanced with acidity and a firm texture, it's very much for food.	87	Alsace	Roger Voss	Trimbach 2012 Gewurztraminer (Alsace)	Gewürztraminer	Trimbach
8	Germany	Savory dried thyme notes accent sunnier flavors of preserved peach in this brisk, off-dry wine. It's fruity and fresh, with an elegant, sprightly footprint.	87	Rheinhessen	Anna Lee C. Iijima	Heinz Eifel 2013 Shine Gewürztraminer (Rheinhe	Gewürztraminer	Heinz Eifel
9	France	This has great depth of flavor with its fresh apple and pear fruits and touch of spice. It's off dry while balanced with acidity and a crisp texture. Drink now.	87	Alsace	Roger Voss	Jean-Baptiste Adam 2012 Les Natures Pinot Gi	Pinot Gris	Jean-Baptiste Adam
10	US	Soft, supple plum envelopes an oaky structure in this Cabernet, supported by 15% Merlot. Coffee and chocolate complete the picture, finishing strong at the er	87	California	Virginie Boone	Kirkland Signature 2011 Mountain Cuvée Cabe	Cabernet Sauvignon	Kirkland Signature
11	France	This is a dry wine, very spicy, with a tight, taut texture and strongly mineral character layered with citrus as well as pepper. It's a food wine with its almost crisp	87	Alsace	Roger Voss	Leon Beyer 2012 Gewurztraminer (Alsace)	Gewürztraminer	Leon Beyer
12	US	Slightly reduced, this wine offers a chalky, tannic backbone to an otherwise juicy explosion of rich black cherry, the whole accented throughout by firm oak and	87	California	Virginie Boone	Louis M. Martini 2012 Cabernet Sauvignon (Ale	Cabernet Sauvignon	Louis M. Martini
13	Italy	This is dominated by oak and oak-driven aromas that include roasted coffee bean, espresso, coconut and vanilla that carry over to the palate, together with pl	87	Sicily & Sardinia	Kerin O'Keefe	Masseria Setteporte 2012 Rosso (Etna)	Nerello Mascalese	Masseria Setteporte
14	US	Building on 150 years and six generations of winemaking tradition, the winery trends toward a leaner style, with the classic California buttercream aroma cut by	87	California	Matt Kettmann	Mirassou 2012 Chardonnay (Central Coast)	Chardonnay	Mirassou
15	Germany	Zesty orange peels and apple notes abound in this sprightly, mineral-toned Riesling. Off dry on the palate, yet racy and lean, it's a refreshing, easy quaffer with	87	Mosel	Anna Lee C. Iijima	Richard Böcking 2013 Devon Riesling (Mosel)	Riesling	Richard Böcking
16	Argentina	Baked plum, molasses, balsamic vinegar and cheesy oak aromas feed into a palate that's braced by a bolt of acidity. A compact set of saucy red-berry and p	87	Other	Michael Schachner	Felix Lavaque 2010 Felix Malbec (Cafayate)	Malbec	Felix Lavaque
17	Argentina	Raw black-cherry aromas are direct and simple but good. This has a juicy feel that thickens over time, with oak character and extract becoming more apparent	87	Mendoza Province	Michael Schachner	Gaucho Andino 2011 Winemaker Selection Mal	Malbec	Gaucho Andino
18	Spain	Desiccated blackberry, leather, charred wood and mint aromas carry the nose on this full-bodied, tannic, heavily oaked Tinto Fino. Flavors of clove and woodsy	87	Northern Spain	Michael Schachner	Pradorey 2010 Vendimia Seleccionada Finca Vi	Tempranillo Blend	Pradorey
19	US	Red fruit aromas pervade on the nose, with cigar box and menthol notes riding in the back. The palate is slightly restrained on entry, but opens up to riper notes	87	Virginia	Alexander Peartree	Quévremont 2012 Meritage (Virginia)	Meritage	Quévremont
20	US	Ripe aromas of dark berries mingle with ample notes of black pepper, toasted vanilla and dusty tobacco. The palate is oak-driven in nature, but notes of tart re	87	Virginia	Alexander Peartree	Quévremont 2012 Vin de Maison Red (Virginia)	Red Blend	Quévremont
21	US	A sleek mix of tart berry, stem and herb, along with a hint of oak and chocolate, this is a fair value in a widely available, drink-now Oregon Pinot. The wine oak-a	87	Oregon	Paul Gregutt	Acrobat 2013 Pinot Noir (Oregon)	Pinot Noir	Acrobat
22	Italy	Delicate aromas recall white flower and citrus. The palate offers passion fruit, lime and white peach with a hint of mineral alongside bright acidity.	87	Sicily & Sardinia	Kerin O'Keefe	Baglio di Pianetto 2007 Ficiligno White (Sicilia)	White Blend	Baglio di Pianetto
23	US	This wine from the Geneseo district offers aromas of sour plums and just enough cigar box to tempt the nose. The flavors are a bit flat at first, then the acidity ar	87	California	Matt Kettmann	Bianchi 2011 Signature Selection Merlot (Paso F	Merlot	Bianchi
24	Italy	Aromas of prune, blackcurrant, toast and oak carry through to the extracted palate, along with flavors of black cherry and roasted coffee beans. Firm but dryin	87	Sicily & Sardinia	Kerin O'Keefe	Canicattì 2009 Aynat Nero d'Avola (Sicilia)	Nero d'Avola	Canicattì
25	US	Oak and earth intermingle around robust aromas of wet forest floor in this vineyard-designated Pinot that hails from a high-elevation site. Small in production, it c	87	California	Virginie Boone	Castello di Amorosa 2011 King Ridge Vineyard	Pinot Noir	Castello di Amorosa
26	Italy	Pretty aromas of yellow flower and stone fruit lead the nose. The bright palate offers yellow apple, apricot, vanilla and delicate notes of lightly toasted oak alongs	87	Sicily & Sardinia	Kerin O'Keefe	Stemmari 2013 Dalla White (Terre Siciliane)	White Blend	Stemmari
27	Italy	Aromas recall ripe dark berry, toast and a whiff of cake spice. The soft, informal palate offers sour cherry, vanilla and a hint of espresso alongside round tannin	87	Sicily & Sardinia	Kerin O'Keefe	Stemmari 2013 Nero d'Avola (Terre Siciliane)	Nero d'Avola	Stemmari
28	Italy	Aromas suggest mature berry, scorched earth, animal, toast and anise. The palate offers ripe black berry, oak, espresso, cocoa and vanilla alongside dusty ta	87	Sicily & Sardinia	Kerin O'Keefe	Terre di Giurfo 2011 Mascaria Barricato (Ceras	Red Blend	Terre di Giurfo
29	US	Clarksburg is becoming a haven for Chenin Blanc in California. This bottling, using fruit sourced from several vineyards in the area, is balanced with just a trace	86	California	Virginie Boone	Clarksburg Wine Company 2010 Chenin Blanc	Chenin Blanc	Clarksburg Wine Company
30	France	Red cherry fruit comes laced with light tannins, giving this bright wine an open, juicy character.	86	Beaujolais	Roger Voss	Domaine de la Madone 2012 Nouveau (Beaujo	Gamay	Domaine de la Madone
31	Italy	Merlot and Nero d'Avola form the base for this easy red wine that would pair with fettuccine and meat sauce or pork roast. The quality of the fruit is clean, bright	86	Sicily & Sardinia		Duca di Salaparuta 2010 Calanica Nero d'Avola	Red Blend	Duca di Salaparuta
32	Italy	Part of the extended Calanica series, this Grillo-Viognier blend shows aromas of honeysuckle and jasmine, backed by touches of cut grass and wild sage. In th	86	Sicily & Sardinia		Duca di Salaparuta 2011 Calanica Grillo-Viogni	White Blend	Duca di Salaparuta

Figura 4.5. *Estructura de la base de datos de comentarios y valoraciones de vinos. Cada fila representa un documento en la base, es decir, un comentario de un experto. Fuente elaboración propia.*

Objetivo del análisis

Desde la Asociación de Enoturismo de España (Wine Tourism Spain) desean conocer cuáles son los principales atributos de los vinos españoles que más son valorados por los expertos, al objeto de llevar a cabo una campaña de publicidad internacional. Para ello los científicos de datos de la asociación seleccionan de la base de comentarios aquellos correspondientes a vinos españoles (6.646 vinos) y aplicando técnicas de minería de datos y de regresión PLS van a tratar de aportar conocimiento para la toma de decisiones por parte de los gestores de la asociación.

Herramienta aplicada

La minería de textos es una herramienta de minería de datos que consiste en examinar colecciones de materiales de texto con el objeto de capturar los temas y conceptos clave y descubrir las relaciones ocultas y las tendencias existentes en los mismos. Busca crear mapas de palabras basados en la 'frecuencia' de ocurrencia y en la 'afinidad' entre las palabras que aparecen en los documentos.

UNIVERSIDAD DE GRANADA
abiertaugr la formación abierta de la UGR
BY NC ND

Entre los beneficios cabe destacar los siguientes [5]:

- Coste de tiempo: permite conseguir grandes economías de escala en comparación a la codificación manual, particularmente cuando se trata de grandes y recurrentes datos que necesitan ser analizados. El proceso es automatizado, aunque cierto grado de intervención humana es necesario para asegurarse de que la exactitud y relevancia funcionan correctamente.
- Consistencia: el análisis es más consistente que la codificación manual, lo que implica que los cambios en los resultados representan la variación real y no producto de un enfoque de codificación diferente. Esto además permite la integración de datos de varias fuentes para hacer posibles comparaciones y vínculos relevantes para una mejor analítica predictiva.
- Simplicidad: los datos no estructurados son notablemente complejos, pero gracias al text-mining, es más fácil filtrar, buscar y hacer referencias cruzadas entre los datos con una sola vista unificada.
- Mejores ideas: el text-mining no es solamente una alternativa más económica que la codificación manual. También una herramienta que incrementa significativamente nuestra habilidad para extraer significados, inteligencia y capacidad de predicción a partir de textos a gran escala.

Resultados

El software empleado para el análisis de datos fue KATE [6], un paquete gratuito de complementos de Excel que incluye una amplia variedad de herramientas de ciencias de datos. Utilizando los módulos WordMap y ExtracText de dicho paquete se llevó a cabo el análisis de minería de textos.

En primer lugar, el software lo que hace es extraer las distintas palabras contenidas en la base de datos de revisiones de vinos por los expertos para posteriormente filtrar ciertas palabras como pronombres, preposiciones, símbolos, números, así como aquellos términos más comúnmente usados en el lenguaje que por ser tan comunes no describen muy bien el contenido de los diferentes textos analizados. Asimismo, la herramienta combina las palabras que comparten la misma raíz (wood, wooden, woody, wooded, woods) en un único término representativo de todas ellas. Así, confecciona un corpus de palabras clave (50, 100 o 150, dependiendo de las demanda del científico) que representarían las opiniones de los expertos en sus valoraciones de los vinos.

Finalmente, la herramienta cuenta esas palabras clave y obtiene la frecuencia con la que aparecen en la base de datos y las relaciones entre ellas. Estas relaciones se basan en la similaridad, es decir, las veces que cada par de palabras aparecen juntas, que representaría la probabilidad de que ambas palabras puedan aparecer juntas en cada documento.

La figura siguiente muestra para cada documento (comentario del experto a un vino) la frecuencia de palabras clave que aparecen en el mismo. Por ejemplo, el comentario 1 incluye 2 veces la palabra “flavors”, 1 vez la palabra “finish”, 1 vez la palabra “aromas”, etc.

UNIVERSIDAD DE GRANADA
abiertaugr la formación abierta de la UGR
CC BY NC ND

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	ID	Points	flavors	finish	aromas	palate	berry	plum	fruit	feels	drink	acidity	nose	fresh	spice	cherry	blackberry	apple	ripe
2	1	98	2	1	1	1	2	0	0	0	1	0	0	0	0	0	1	0	1
3	2	97	1	0	1	1	0	0	1	0	1	0	0	0	0	0	1	0	0
4	3	97	1	1	1	0	0	1	0	0	1	1	0	0	0	0	2	0	0
5	4	97	1	1	1	0	0	1	0	0	1	0	1	0	0	1	0	0	0
6	5	97	1	0	1	0	0	0	1	0	1	1	0	0	0	1	1	0	1
7	6	97	1	0	0	1	2	0	1	0	1	1	1	0	2	0	0	0	0
8	7	96	1	0	1	0	0	0	0	1	1	0	0	0	1	0	2	0	0
9	8	96	1	1	1	1	0	1	0	0	1	0	0	0	1	1	0	0	0
10	9	96	2	1	1	1	0	0	0	0	1	0	0	0	0	0	2	0	0
11	10	96	1	1	1	0	0	0	0	0	1	0	0	0	0	0	2	0	1
12	11	96	1	1	1	0	1	1	1	0	0	0	0	0	0	0	0	0	0
13	12	96	2	1	1	0	1	0	1	0	1	0	0	0	1	1	2	0	0
14	13	96	1	0	0	1	1	0	0	0	0	0	1	0	0	0	0	0	1
15	14	96	1	1	0	1	1	0	0	0	1	0	1	0	0	0	1	0	1
16	15	96	1	1	1	1	0	1	0	0	2	0	0	0	0	1	0	0	0
17	16	96	2	1	1	0	0	0	0	0	1	0	1	0	2	0	1	0	0
18	17	96	1	1	1	0	0	0	1	1	1	0	0	0	0	0	0	0	0
19	18	96	1	1	1	1	0	0	0	0	1	1	0	0	0	0	0	0	0
20	19	96	1	1	1	1	2	1	2	0	1	0	0	0	2	0	0	0	0
21	20	96	1	1	1	1	1	0	0	0	1	0	0	0	1	0	1	0	0
22	21	96	1	1	1	1	1	0	1	0	1	0	0	0	0	0	0	0	0
23	22	96	1	0	1	1	0	0	2	0	0	0	0	0	1	0	0	0	0
24	23	95	1	1	1	0	0	0	0	0	0	0	0	0	0	0	2	0	1
25	24	95	1	1	1	0	0	0	1	1	1	0	1	1	2	0	1	0	0
26	25	95	1	1	1	1	0	1	1	0	1	0	1	0	1	1	0	0	0
27	26	95	1	0	1	0	0	0	0	0	1	0	1	0	0	1	2	0	1
28	27	95	1	0	1	0	2	0	0	0	1	0	0	0	1	1	0	0	0
29	28	95	1	1	1	1	0	0	1	0	1	0	0	0	1	1	0	0	0
30	29	95	1	1	0	1	1	0	1	0	0	0	0	0	0	0	0	0	1
31	30	95	1	1	1	1	0	0	0	1	1	0	0	1	1	1	2	0	1
32	31	95	1	1	1	0	0	0	0	1	0	1	1	0	0	0	0	0	0
33	32	95	1	1	0	1	1	0	1	0	0	0	1	0	0	1	1	0	0
34	33	95	1	1	1	1	0	1	0	0	0	0	0	0	0	0	2	0	0
35	34	95	1	1	1	1	1	0	2	0	1	0	1	0	0	1	0	0	0

Figura 6. *Matriz de frecuencia de palabras clave para cada documento. Fuente elaboración propia*

La figura 4.7 recoge la matriz de similaridad de las 100 palabras clave seleccionadas. Un análisis de la matriz muestra los pares de palabras que son más probables que aparezcan juntos en los documentos. Así, “flavors” y “finish” son nombradas 4764 veces juntas por los expertos en sus comentarios, o “aromas” y “palate” aparecen nombradas conjuntamente en 2525 veces. El análisis detenido de esta matriz ya nos estaría describiendo de una manera rápida los atributos más destacables que los expertos consideran que tienen los vinos españoles contenidos en la base de datos, uno de los objetivos que buscaba la Asociación de Enoturismo de España.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1		flavors	finish	aromas	palate	berry	plum	fruit	feels	drink	acidity	nose	fresh	spice	cherry	blackberry	apple	ripe
2	flavors	0	4764	4701	2894	1931	1819	1700	1940	1631	1401	1194	1078	1021	974	943	843	809
3	finish	4764	0	4171	2563	1647	1577	1503	1698	1435	1179	978	921	918	816	816	721	701
4	aromas	4701	4171	0	2525	1711	1593	1526	1679	1425	1191	644	945	900	866	829	715	695
5	palate	2894	2563	2525	0	967	903	957	545	892	760	743	569	549	538	493	468	413
6	berry	1931	1647	1711	967	0	964	777	732	667	401	340	318	502	374	286	10	388
7	plum	1819	1577	1593	903	964	0	441	634	587	485	324	392	452	626	320	19	300
8	fruit	1700	1503	1526	957	777	441	0	605	542	394	362	300	379	258	239	144	276
9	feels	1940	1698	1679	545	732	634	605	0	430	412	363	380	320	294	236	280	256
10	drink	1631	1435	1425	892	667	587	542	430	0	371	306	253	462	319	539	148	436
11	acidity	1401	1179	1191	760	401	485	394	412	371	0	298	308	209	261	126	275	140
12	nose	1194	978	644	743	340	324	362	363	306	298	0	212	197	190	167	187	170
13	fresh	1078	921	945	569	318	392	300	380	253	308	212	0	186	242	64	236	77
14	spice	1021	918	900	549	502	452	379	320	462	209	197	186	0	242	266	87	210
15	cherry	974	816	866	538	374	626	258	294	319	261	190	242	242	0	140	12	121
16	blackberry	943	816	829	493	286	320	239	236	539	126	167	64	266	140	0	0	337
17	apple	843	721	715	468	10	19	144	280	148	275	187	236	87	12	0	0	28
18	ripe	809	701	695	413	388	300	276	256	436	140	170	77	210	121	337	28	0
19	citrus	731	639	624	419	32	35	239	238	125	212	152	232	33	34	3	250	31
20	tannins	777	686	672	394	413	345	276	222	362	158	142	65	183	159	259	1	204
21	black	720	627	636	385	273	321	464	205	344	119	124	79	201	258	278	0	211
22	good	739	572	605	402	282	245	243	256	157	211	188	152	162	159	94	128	99
23	herbal	679	582	573	337	356	320	226	276	148	132	115	93	130	115	111	5	73
24	tastes	494	672	606	433	241	225	271	276	180	138	155	131	135	102	95	101	95
25	blend	731	632	600	390	283	252	215	232	219	184	175	129	153	140	127	83	137
26	vanilla	689	584	607	369	283	254	257	229	306	197	148	120	190	155	145	75	138
27	raspberry	695	558	587	377	198	463	168	223	193	214	150	195	148	279	47	7	59
28	baked	668	587	587	341	366	260	259	220	312	126	124	52	234	102	181	75	182
29	earthy	661	606	614	323	409	374	219	246	305	137	89	88	176	154	148	2	139
30	tannic	620	557	539	298	312	265	218	190	252	61	123	51	138	104	237	1	118
31	chocolate	606	525	515	308	295	205	194	172	319	80	130	73	191	105	262	3	172
32	sweet	519	409	381	318	173	127	156	170	114	118	168	90	65	96	69	80	105
33	green	526	446	447	277	95	69	162	194	71	136	140	112	27	26	23	215	17
34	full	582	523	493	305	261	219	182	198	254	131	121	67	139	111	171	49	137
35	touch	546	509	488	279	221	197	168	211	140	132	121	112	92	122	73	81	84
36	melon	489	412	408	264	3	3	100	164	94	163	106	119	27	3	0	217	25
37	oaky	494	465	439	252	254	221	178	166	201	122	93	74	145	93	111	24	100
38	mouth	516	445	405	30	179	163	163	87	166	151	127	77	92	92	96	60	91
39	herbs	490	430	411	252	203	196	151	194	187	102	100	110	134	83	103	52	67
40	smells	452	400	168	260	135	126	159	179	106	95	89	75	79	58	54	50	43
41	dark	442	388	374	247	205	178	176	118	201	66	103	39	124	87	226	1	112

Figura 7. *Matriz de similaridad de palabras clave. Fuente elaboración propia*

UNIVERSIDAD DE GRANADA
abiertaugr la formación abierta de la UGR
BY NC ND

Llegados a este punto, un análisis adicional que se podría llevar a cabo sería examinar la relación que existe entre las palabras clave identificadas y las valoraciones cuantitativas que los expertos han dado a cada vino, al objeto de analizar qué atributos se identifican preferiblemente con valoraciones altas y cuáles con valoraciones bajas. Para ello, se llevó a cabo un análisis de regresión considerando como variable dependiente o explicativa la valoración cuantitativa del experto y como variables independientes (predictores) las 100 palabras clave identificadas. En lugar de utilizar un método de regresión convencional se optó por la regresión PLS que permite reducir de manera simultánea el espacio de los 100 predictores en unos cuantos componentes (en nuestro caso a 2) maximizando el poder explicativo sobre la variable dependiente.

Los resultados mostrados en la figura siguiente señalan un R^2 del 55,1% que informa sobre el poder explicativo del modelo con dos dimensiones. La columna C recogería los coeficientes estandarizados para cada palabra clave. En color verde se señalan las palabras clave (predictores) con un impacto positivo mayor en la valoración del vino (drink, spice, blackberry, ripe, chocolate, minerally) y en rojo las palabras con un impacto más negativo en dicha valoración (feels, herbal, green, smells, bitter). Es decir, cuando los expertos citan en sus revisiones las palabras señaladas en verde la probabilidad de que otorguen una valoración positiva al vino será muy alta y lo contrario cuando citen palabras en rojo.

	A	B	C	D	E	F	G	H
3		Results from the PLS regression	Y Rsq	X Rsq				
4			55,1%	07,1%				
5	Predictor	Regression coefficient	Standardized Coefficient	Predictor	Score coeff. 1	Score coeff. 2	Loadings 1	Loadings 2
6	flavors	0,184	0,027	flavors	0,029	-0,006	0,088	0,010
7	finish	0,411	0,056	finish	0,040	0,009	0,085	0,053
8	aromas	0,280	0,039	aromas	0,036	-0,002	0,097	0,036
9	palate	0,124	0,021	palate	0,011	0,008	0,013	0,062
10	berry	0,085	0,017	berry	0,033	-0,022	0,132	-0,201
11	plum	0,066	0,013	plum	0,027	-0,019	0,109	-0,163
12	fruit	0,006	0,001	fruit	0,007	-0,007	0,031	-0,057
13	feels	-0,246	-0,041	feels	-0,028	-0,009	-0,054	-0,067
14	drink	1,596	0,232	drink	0,166	0,040	0,350	0,136
15	acidity	0,083	0,013	acidity	-0,003	0,017	-0,043	0,088
16	nose	-0,075	-0,010	nose	-0,015	0,008	-0,056	-0,003
17	fresh	0,307	0,042	fresh	0,005	0,037	-0,065	0,188
18	spice	0,562	0,081	spice	0,072	-0,003	0,194	-0,017
19	cherry	-0,008	-0,001	cherry	0,010	-0,013	0,055	-0,090
20	blackberry	0,802	0,117	blackberry	0,111	-0,012	0,315	-0,025
21	apple	0,263	0,041	apple	-0,009	0,051	-0,128	0,303
22	ripe	0,658	0,085	ripe	0,080	-0,007	0,224	-0,037
23	citrus	0,099	0,013	citrus	-0,026	0,043	-0,157	0,234
24	tannins	0,384	0,046	tannins	0,053	-0,015	0,168	-0,120
25	black	0,488	0,063	black	0,067	-0,014	0,203	-0,069
26	good	0,118	0,014	good	0,003	0,011	-0,014	0,043
27	herbal	-0,321	-0,040	herbal	-0,019	-0,018	-0,011	-0,171
28	tastes	-0,278	-0,031	tastes	-0,024	-0,004	-0,055	-0,055
29	blend	0,090	0,010	blend	0,009	-0,001	0,026	-0,030
30	vanilla	0,393	0,048	vanilla	0,044	-0,003	0,120	0,020
31	raspberry	-0,150	-0,018	raspberry	-0,011	-0,005	-0,018	-0,074
32	baked	0,196	0,023	baked	0,039	-0,022	0,147	-0,096
33	earthy	0,177	0,020	earthy	0,033	-0,018	0,122	-0,115
34	tannic	0,209	0,024	tannic	0,033	-0,015	0,119	-0,129
35	chocolate	0,805	0,085	chocolate	0,077	-0,005	0,210	-0,029
36	sweet	-0,197	-0,022	sweet	-0,020	0,001	-0,053	-0,039
37	green	-0,341	-0,039	green	-0,041	0,009	-0,126	0,074
38	full	0,516	0,052	full	0,048	-0,004	0,133	-0,011
39	touch	0,282	0,028	touch	0,010	0,016	-0,007	0,042
40	melon	0,206	0,021	melon	-0,015	0,038	-0,117	0,223
41	oaky	0,157	0,015	oaky	0,026	-0,014	0,096	-0,076
42	mouth	-0,026	-0,002	mouth	0,004	-0,007	0,023	-0,029
43	herbs	0,320	0,030	herbs	0,023	0,003	0,055	0,004
44	smells	-0,595	-0,053	smells	-0,037	-0,010	-0,075	-0,089
45	dark	0,413	0,039	dark	0,050	-0,019	0,168	-0,066
46	minerally	0,906	0,086	minerally	0,040	0,040	0,020	0,247
47	toasty	0,858	0,080	toasty	0,073	-0,005	0,200	0,011
48	bitter	-0,630	-0,057	bitter	-0,048	-0,002	-0,121	0,003
49	peach	0,253	0,024	peach	-0,010	0,036	-0,100	0,235

Figura 8. *Resultados del análisis de regresión PLS. Fuente elaboración propia*

UNIVERSIDAD DE GRANADA
abiertaugr la formación abierta de la UGR
CC BY NC ND

Podemos representar las valoraciones cuantitativas otorgadas a cada uno de los vinos contenidos en la base de datos en la solución de dos dimensiones proporcionada por el análisis de regresión PLS. Se observa gráficamente en la figura 4.9 como en la parte derecha se situarían los vinos con una valoración más alta (media: 92,18 puntos), mientras que en la parte izquierda se situarían los vinos peor valorados (media: 84,50 puntos).

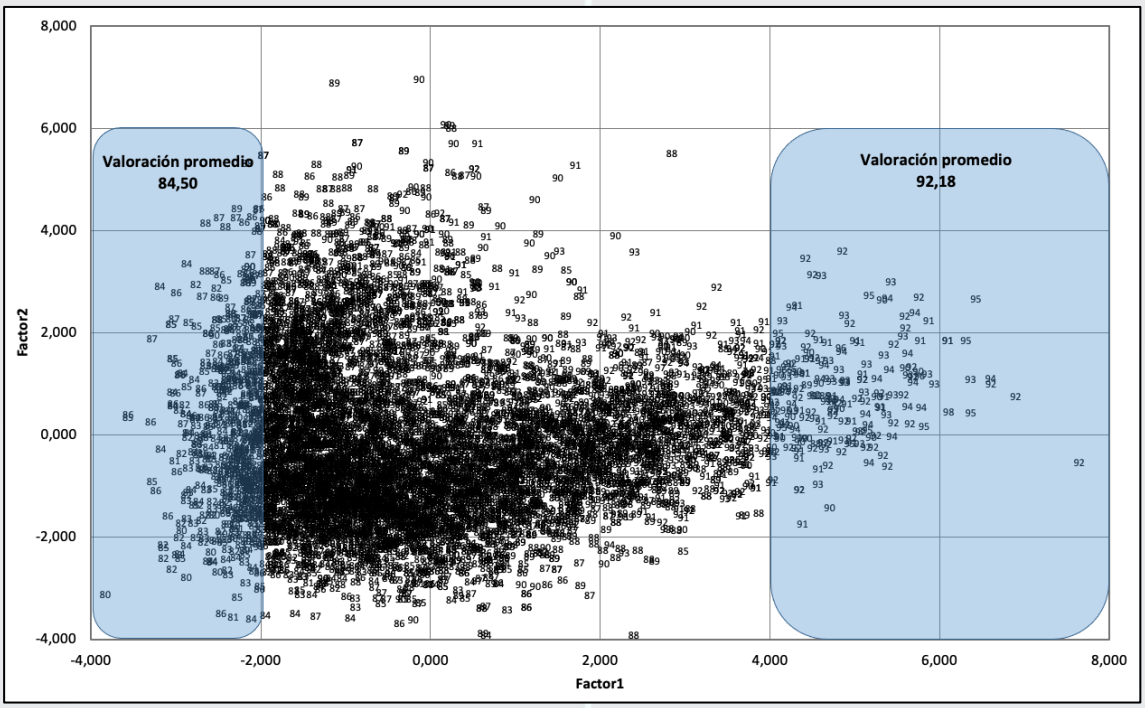

Figura 9. *Representación gráfica de los 2 componentes y las valoraciones cuantitativas al vino. Fuente elaboración propia.*

Si en ese mismo plano de las dos dimensiones representamos el corpus de palabras clave (figura 10), podemos concluir qué palabras están asociadas claramente a altas valoraciones de los expertos (drink, blackberry, ripe, chocolate, cassis, spice, toasty, tannins…) y qué otras palabras se asocian claramente a bajas puntuaciones (bitter, candied, smells, tastes, basic, herbal, tomato, mouthfeel…).

En definitiva, con la aplicación de la minería de textos en combinación con alguna que otra herramienta estadística como la regresión hemos podido extraer un rico conocimiento para la toma de decisiones de una base extensa de valoraciones de expertos, que de otra forma hubiera sido imposible analizar.

UNIVERSIDAD DE GRANADA
abiertaugr
la formación abierta de la UGR
BY NC ND

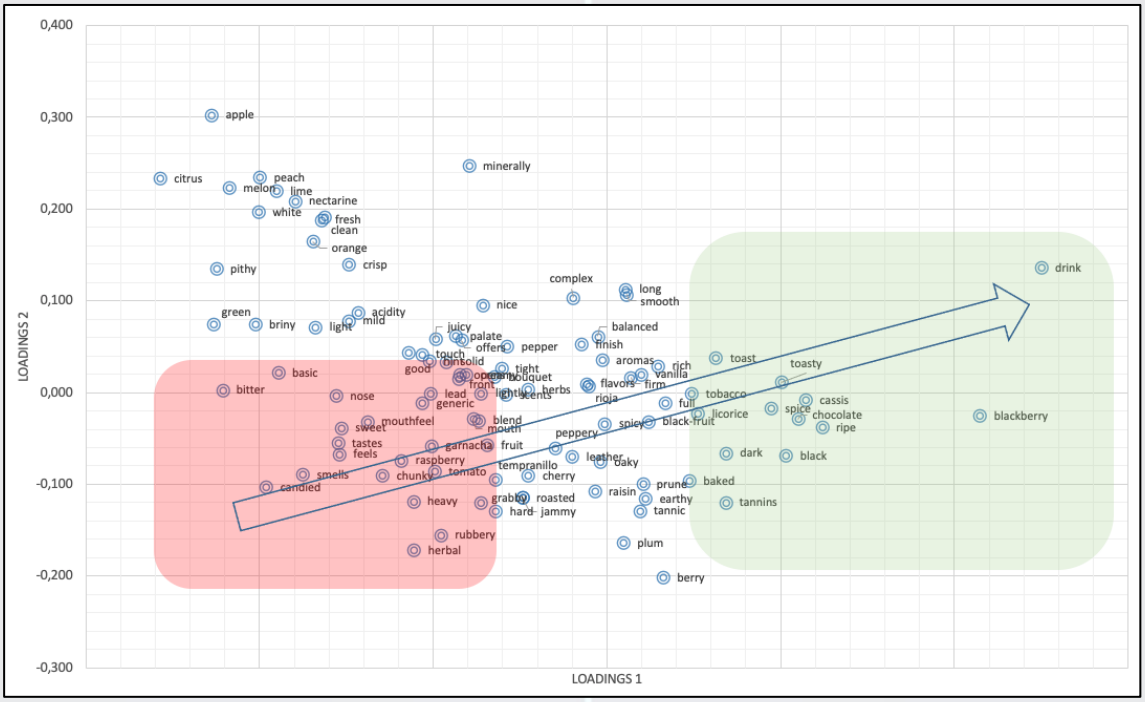

Figura 10. *Representación gráfica de los 2 componentes y las palabras clave. Fuente elaboración propia.*

Aplicación de la herramienta de clustering

Descripción de la base de datos

Contamos con una base de datos de 50.000 clientes de una empresa de los que conocemos una serie de variables relacionadas con sus estilos de vida (aficiones y comportamientos durante su tiempo de ocio): asistir al teatro, asistir a fiestas, leer frecuentemente revistas y libros, salir frecuentemente con los amigos, leer periódicos, salir de vacaciones, visitar a familiares, escuchar la radio, asistir a conciertos en vivo, cuidar del jardín en el hogar, ver la televisión, salir a bares con amigos, escuchar música, viajar con la familia, asistir a actos religiosos.

Cada sujeto tiene una puntuación de 1 a 4 en cada una de esas variables, representando una escala que va de 1: poca probabilidad de que tenga ese estilo de vida a 4: alta probabilidad de que tenga este estilo de vida.

UNIVERSIDAD DE GRANADA
abiertaugr la formación abierta de la UGR
CC BY NC ND

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	ID	ATTEND THEATER	PARTIES OR SOCIAL FUNCTIONS	READING: MAGAZINES	READING: BOOKS	GO OUT WITH FRIENDS	NEWS PAPERS	TRAVEL ON VACATION	VISIT RELATIVES	LISTEN TO RADIO	LIVE MUSIC SHOWS	GARDEN	WATCH TV	GO TO BARS	LISTEN TO MUSIC	OUT WITH THE FAMILY	ATTEND RELIGIOUS SERVICES
2	Subject 1	2	3	3	3	3	2	2	3	3	2	1	3	1	2	3	2
3	Subject 2	2	2	2	2	2	2	4	2	2	2	1	3	2	2	2	2
4	Subject 3	1	2	2	1	3	4	2	2	3	1	2	4	3	3	2	2
5	Subject 4	1	2	2	1	1	2	2	2	4	2	3	4	1	1	3	1
6	Subject 5	1	2	3	2	2	3	2	2	3	1	1	3	2	3	3	4
7	Subject 6	2	2	3	2	2	3	3	2	3	2	1	3	1	3	3	2
8	Subject 7	1	2	2	2	2	2	2	1	4	2	3	3	2	3	2	2
9	Subject 8	2	2	2	2	2	2	2	2	3	2	1	4	2	3	2	1
10	Subject 9	2	2	2	2	4	4	3	2	4	4	1	4	4	4	1	2
11	Subject 10	1	2	1	2	2	2	2	3	3	1	2	3	2	2	3	2
12	Subject 11	1	1	2	1	1	2	1	1	4	2	2	4	1	4	2	1
13	Subject 12	2	3	3	3	2	3	3	2	3	2	1	4	3	4	2	2
14	Subject 13	3	2	1	4	3	2	3	2	4	4	3	2	3	3	3	3
15	Subject 14	1	2	2	2	2	2	2	4	4	1	3	2	2	2	1	4
16	Subject 15	2	3	3	4	3	2	3	4	4	2	3	4	3	2	3	3
17	Subject 16	2	3	2	2	2	1	2	2	2	2	1	3	2	3	1	2
18	Subject 17	2	3	4	2	3	3	3	3	4	1	3	4	1	3	4	4
19	Subject 18	2	3	3	2	4	3	2	3	4	2	3	4	1	3	4	3
20	Subject 19	2	3	3	3	3	3	2	2	4	2	2	4	2	3	2	1
21	Subject 20	2	2	3	2	2	3	3	3	3	2	2	3	2	2	3	2
22	Subject 21	2	2	3	4	3	2	3	2	3	2	2	3	2	3	3	2
23	Subject 22	1	3	2	4	3	2	2	2	3	4	4	2	4	4	1	1
24	Subject 23	2	1	2	1	2	2	2	2	3	2	3	4	1	3	3	1
25	Subject 24	1	3	1	3	1	2	3	3	2	2	4	4	3	1	2	2
26	Subject 25	2	2	2	2	2	2	2	2	2	2	3	3	2	2	3	2
27	Subject 26	1	2	3	3	2	3	2	3	4	1	2	4	1	2	3	1
28	Subject 27	2	2	4	3	3	2	4	2	4	2	2	4	2	3	2	1
29	Subject 28	2	2	3	4	2	2	2	2	3	2	3	3	2	3	4	1
30	Subject 29	1	4	3	2	4	2	3	3	3	2	2	3	3	3	3	2
31	Subject 30	1	2	3	3	2	3	2	3	3	1	3	3	1	3	3	2
32	Subject 31	1	2	2	2	1	2	1	2	3	1	2	4	2	1	2	1
33	Subject 32	2	2	3	3	1	3	2	2	3	1	1	3	1	2	2	2
34	Subject 33	1	2	2	3	1	3	2	3	4	1	3	4	1	2	3	3
35	Subject 34	1	2	1	2	1	2	2	2	3	2	1	4	4	4	3	2
36	Subject 35	1	4	4	2	2	3	4	2	3	2	1	4	1	3	2	4
37	Subject 36	1	2	2	2	2	2	2	4	4	1	1	4	2	4	4	2
38	Subject 37	2	2	3	4	3	4	4	4	2	1	1	2	2	3	3	1

Figura 11. *Estructura de la base de datos estructurada. Cada fila representa a cada cliente y cada columna cada una de las variables consideradas. Fuente elaboración propia.*

Objetivo del análisis

La empresa quiere segmentar el mercado de sus clientes en varios grupos homogéneos en cuanto a sus estilos de vida de cara a plantear estrategias comerciales diferentes adaptadas a cada perfil de cliente.

Herramienta aplicada

Las herramientas de clustering permiten agrupar o segmentar casos a través del uso de procedimientos jerárquicos usando reglas, probabilidades y estadísticos. Dentro de este tipo destacan el análisis cluster (k-means) y el análisis de clases latentes. Ambas técnicas agrupan casos individuales en grupos relativamente homogéneos, al mismo tiempo que identifican las características de esos segmentos. La principal diferencia entre ambas técnicas es que k-means utiliza la distancia métrica entre los individuos para agruparlos, mientras que clases latentes utiliza la probabilidad de pertenencia de un individuo a un grupo mediante un algoritmo iterativo (algoritmo E-M). Esto tienen importantes ventajas respecto a k-means en el sentido de que permite utilizar tanto variables de tipo cuantitativo como cualitativo, además de que la existencia de valores perdidos (missing data) en la base no representa un problema. En este ejemplo trabajaremos con el análisis de clases latentes.

UNIVERSIDAD DE GRANADA
abiertaugr la formación abierta de la UGR
CC BY NC ND

Resultados

De nuevo utilizamos el software KATE [6] a través del módulo Latent Class Analysis.

Una vez aplicada la herramienta y seleccionada una solución de tres clases o segmentos por ser la que presenta un valor de ajuste predictivo más adecuado de acuerdo al criterio estadístico CAIC, la figura 4.12 muestra los resultados. La primera clase o segmento incluiría al 34,4% de los clientes, la segunda al 39,9% y la tercera al 25,7%. Para cada variable se presentan las probabilidades asociadas a cada clase, en azul se señalan las más altas y en rojo las más bajas, y ello permite examinar las diferencias en los estilos de vida entre esos tres segmentos e interpretar los resultados.

Así, los individuos que forman parte del segmento 1 presentan una alta probabilidad (muy superior al resto de los segmentos) de ir al teatro, acudir a eventos sociales, salir con amigos, ir de vacaciones, escuchar música en vivo, ir a bares y escuchar música, y menos probabilidad de cuidar el jardín en el hogar o acudir a actividades de carácter religioso. Es decir, se trataría de personas con un estilo de vida muy activo y social. En cambio, los clientes del segmento 2 se caracterizarían por tener una probabilidad superior al resto en atender el jardín en el hogar, salir con la familia y atender actividades religiosas. Por tanto, son un grupo de personas totalmente diferentes a las del segmento 1 con un estilo de vida mucho más familiar y tradicional. Finalmente, el grupo 3 estaría formado por personas cuya probabilidad de realizar la mayor parte de esas actividades estaría por debajo del resto de grupos y sería similar en lo relativo a visitar a familiares, escuchar la radio, ver la televisión o acudir a actos religiosos.

	A	B	C	D	E
1			Results from the Latent Class Analysis		
2		LogLikelihood=	-7287,5	Overlap=	29,7%
3		CAIC=	14874,8		
4			Response Probabilities within Classes		
5		Class Sizes=>	34,4%	39,9%	25,7%
6	Item	Response	Class 1	Class 2	Class 3
7	ATTEND THEATER	Ordered	25,3%	18,0%	6,3%
8					
9	PARTIES OR SOCIAL FUNCTIONS	Ordered	53,7%	38,5%	26,2%
10					
11	READING: MAGAZINES	Ordered	55,3%	53,4%	26,6%
12					
13	READING: BOOKS	Ordered	51,3%	50,9%	28,0%
14					
15	GO OUT WITH FRIENDS	Ordered	74,5%	39,2%	28,5%
16					
17	NEWS PAPERS	Ordered	51,6%	56,8%	35,0%
18					
19	TRAVEL ON VACATION	Ordered	63,3%	50,8%	34,6%
20					
21	VISIT RELATIVES	Ordered	42,2%	56,6%	39,9%
22					
23	LISTEN TO RADIO	Ordered	82,9%	76,1%	60,8%
24					
25	LIVE MUSIC SHOWS	Ordered	43,1%	20,9%	6,5%
26					
27	GARDEN	Ordered	22,1%	51,0%	20,2%
28					
29	WATCH TV	Ordered	67,3%	77,8%	66,7%
30					
31	GO TO BARS	Ordered	59,0%	20,5%	14,9%
32					
33	LISTEN TO MUSIC	Ordered	73,6%	58,5%	42,8%
34					
35	OUT WITH THE FAMILY	Ordered	45,5%	67,5%	42,2%
36					
37	ATTEND RELIGIOUS SERVICES	Ordered	25,9%	42,4%	30,7%

Figura 12. *Resultados de la agrupación en tres clases o segmentos. Fuente elaboración propia*

UNIVERSIDAD DE GRANADA
abiertaugr la formación abierta de la UGR
CC BY NC ND

La herramienta permite también asignar a cada uno de los clientes que forman parte de la base a una de las tres clases o segmentos, otorgándole además una probabilidad de pertenencia. Como se puede observar en la figura 4.13 el sujeto 1 ha sido asignado a la clase 2 por tener una probabilidad de pertenencia del 74,3%, mientras que el sujeto 9 es claramente incluido en la clase 1 con un 100% de probabilidad.

Membership Probabilities				
Subject	CLASS	Class 1	Class 2	Class 3
Subject 1	2	21,7%	74,3%	4,1%
Subject 2	3	27,9%	19,3%	52,7%
Subject 3	2	33,1%	35,9%	30,9%
Subject 4	3	0,0%	34,6%	65,3%
Subject 5	2	1,8%	63,6%	34,6%
Subject 6	2	13,1%	79,6%	7,3%
Subject 7	2	12,1%	58,9%	29,0%
Subject 8	3	35,3%	18,4%	46,3%
Subject 9	1	100,0%	0,0%	0,0%
Subject 10	3	0,1%	10,4%	89,4%
Subject 11	3	0,1%	4,3%	95,6%
Subject 12	1	96,5%	3,5%	0,0%
Subject 13	1	96,7%	3,3%	0,0%
Subject 14	2	0,3%	54,2%	45,5%
Subject 15	2	21,5%	78,5%	0,0%
Subject 16	3	20,4%	3,5%	76,1%
Subject 17	2	0,2%	99,8%	0,0%
Subject 18	2	2,6%	97,4%	0,0%
Subject 19	1	85,6%	14,4%	0,1%
Subject 20	2	6,5%	92,1%	1,3%
Subject 21	1	58,9%	41,0%	0,1%
Subject 22	1	99,9%	0,1%	0,0%
Subject 23	2	0,4%	73,1%	26,5%
Subject 24	2	0,4%	90,5%	9,1%
Subject 25	2	1,3%	81,3%	17,4%
Subject 26	2	0,3%	91,6%	8,1%
Subject 27	1	92,2%	7,8%	0,0%
Subject 28	2	4,2%	95,7%	0,1%
Subject 29	1	97,7%	2,2%	0,0%
Subject 30	2	0,1%	97,5%	2,4%

Figura 13. *Probabilidad de pertenencia de cada individuo a cada una de las tres clases. Fuente elaboración propia.*

En definitiva, esta herramienta de clustering es especialmente útil para poder identificar segmentos de clientes en base a diferentes características y poder predecir el segmento de pertenencia de cada individuo con una determinada probabilidad.

Aplicación de la herramienta de análisis a datos longitudinales a lo largo del tiempo (Dynamic Factor Analysis)

Descripción de la base de datos

Contamos con una base de datos que incluye la evolución mensual desde el año 2012 hasta la actualidad del interés de la población en veintidós marcas de automóviles. Ese interés es medido de acuerdo a los datos que proporciona Google Trends [7] sobre tendencias de búsqueda en Internet para cada una de las marcas al mes. Los números que recoge la base reflejan el interés de búsqueda en relación con el valor máximo de un gráfico en una región y un periodo determinados. Un valor de

UNIVERSIDAD DE GRANADA
abiertaugr la formación abierta de la UGR
CC BY NC ND

100 indica la popularidad máxima de un término, mientras que 50 indican que un término es la mitad de popular en relación con el valor máximo.

Para cada marca se recoge ese valor de interés a nivel mundial mes a mes.

La figura 14 recoge un ejemplo de extracción de la información sobre el interés de la marca Audi en Google Trends, pudiendo delimitar el área geográfica de consulta y el periodo de búsqueda.

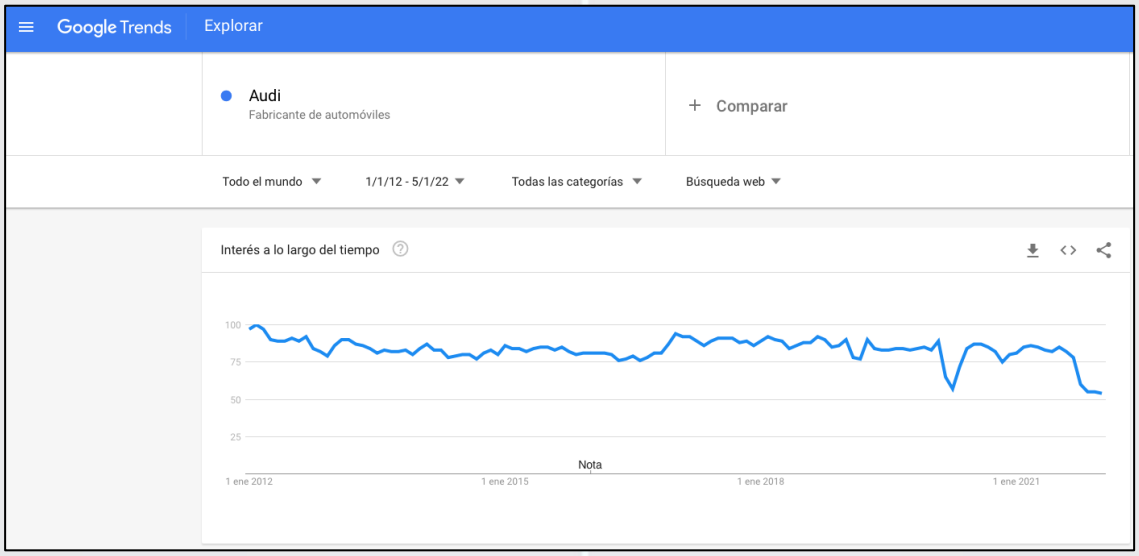

Figura 14. Ejemplo de extracción de la información sobre interés de la marca de automóvil en Google Trends. Fuente elaboración propia.

La figura 15 muestra la base de datos estructurada con la que contaríamos en este ejemplo una vez extraída la información para cada marca de automóviles en Google Trends. Cada fila representaría el periodo temporal, un mes en este caso, y cada columna una marca.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
1	Semana	Alfa Romeo	Audi	Chevrolet	Citroën	Dacia	Fiat	Honda	Hyundai	Kia	Lancia	Mazda	Mercedes	Mitsubishi	Nissan	Opel	Peugeot	Renault	SEAT	Škoda	Toyota	Volkswagen	Volvo
2	12/2/12	86	73	67	86	26	88	81	45	51	91	57	59	10	58	100	70	80	59	58	32	20	30
3	19/2/12	93	66	69	88	24	90	85	64	51	68	61	60	9	77	89	70	74	49	40	31	19	35
4	26/2/12	69	65	64	79	20	69	79	68	47	80	53	57	18	66	82	66	68	54	33	28	17	31
5	4/3/12	77	62	73	100	37	76	94	48	41	69	54	66	9	68	85	65	74	59	39	29	19	44
6	11/3/12	66	65	84	90	37	75	92	58	41	76	51	63	13	90	91	66	79	55	34	28	19	30
7	18/3/12	58	64	69	86	26	81	85	42	45	70	51	62	8	69	86	75	81	50	28	27	18	40
8	25/3/12	55	62	53	71	18	69	86	37	39	72	46	54	10	64	74	67	68	46	26	25	16	31
9	1/4/12	57	62	60	62	17	62	83	30	37	64	52	56	10	46	67	62	59	52	28	23	16	29
10	8/4/12	53	60	59	77	29	69	80	45	43	74	53	56	12	68	76	63	74	46	29	25	16	36
11	15/4/12	61	59	55	80	31	72	80	43	55	64	41	54	7	56	75	73	100	48	37	26	16	32
12	22/4/12	58	56	56	74	26	78	79	38	45	74	47	54	13	50	67	71	82	43	42	26	16	36
13	29/4/12	53	58	61	78	22	63	81	49	46	89	53	55	11	60	64	65	70	42	43	27	17	20
14	6/5/12	62	53	58	71	19	64	79	32	47	80	47	55	13	61	74	62	65	42	31	28	16	33
15	13/5/12	60	63	65	84	21	69	92	43	43	75	51	56	28	72	73	62	65	46	32	24	16	30
16	20/5/12	63	67	60	87	58	73	87	48	53	91	46	57	24	69	73	72	63	43	34	27	16	30
17	27/5/12	62	57	53	76	42	59	88	44	49	72	42	56	32	51	69	75	63	48	43	25	16	41
18	3/6/12	60	57	53	74	37	64	87	54	48	82	46	55	29	65	74	75	64	40	43	26	16	25
19	10/6/12	57	53	50	78	27	66	79	64	51	74	43	56	21	72	78	72	60	43	37	27	16	35
20	17/6/12	58	58	52	73	24	65	83	58	46	74	47	90	22	65	74	79	67	40	36	26	16	49
21	24/6/12	44	53	55	72	24	75	84	77	42	74	50	78	54	56	81	67	64	49	45	27	16	24
22	1/7/12	54	52	67	78	29	73	85	68	37	67	45	68	29	57	73	68	69	46	45	27	17	44
23	8/7/12	58	58	65	76	25	66	89	59	39	69	49	73	18	72	85	80	67	51	55	26	17	22
24	15/7/12	61	59	58	76	25	68	90	46	57	89	46	63	21	55	83	78	64	50	66	26	18	31
25	22/7/12	58	59	55	76	22	70	91	50	42	75	47	60	31	70	79	84	68	51	49	27	17	52
26	29/7/12	60	58	59	80	18	63	85	35	58	48	53	55	28	60	81	78	67	48	40	26	18	43
27	5/8/12	57	56	58	79	22	66	86	55	46	56	48	58	19	75	85	83	66	54	34	36	19	38
28	12/8/12	66	60	70	84	22	69	86	51	66	71	51	61	19	70	80	71	69	50	35	35	17	31
29	19/8/12	62	60	59	81	22	69	94	45	59	76	44	60	29	67	84	74	71	54	49	35	20	36
30	26/8/12	60	57	52	76	20	65	88	41	54	64	50	59	25	57	77	64	59	56	39	28	18	36
31	2/9/12	50	53	48	72	19	62	85	37	25	63	42	58	10	56	79	63	60	45	35	24	17	36
32	9/9/12	62	60	49	65	24	62	82	44	41	61	40	59	9	57	72	69	65	50	31	27	17	32
33	16/9/12	58	62	43	72	33	70	86	49	46	59	42	63	11	52	68	64	65	46	29	27	15	29
34	23/9/12	56	60	44	68	28	69	77	47	39	73	41	73	11	49	72	68	66	52	28	27	15	39
35	30/9/12	61	53	65	77	100	71	81	51	51	59	52	71	9	64	86	73	85	63	35	32	17	27

Figura 15. *Estructura de la base de datos de interés de búsqueda para cada marca a lo largo del tiempo. Fuente elaboración propia.*

UNIVERSIDAD DE GRANADA
abiertaugr
la formación abierta de la UGR
BY NC ND

Objetivo del análisis

La dirección de marketing de cualquiera de las marcas de automóviles consideradas estará interesada en examinar la evolución del interés que despierta cada marca en el mercado a lo largo del tiempo, lo cual permitirá ver las que están perdiendo interés de manera progresiva, que lo están ganando o que están estancadas. Asimismo, será interesante examinar grupos de marca que tienen una tendencia a lo largo del tiempo creciente o decreciente.

Herramienta aplicada

Es muy frecuente encontrarse con datos de carácter longitudinal procedentes de una misma fuente pero a lo largo del tiempo. Para los que tradicionalmente se ha aplicado el análisis de series temporales. Una serie temporal es una secuencia de N observaciones (datos) ordenadas y equidistantes cronológicamente sobre una variable o característica objeto de estudio. En esta técnica se busca examinar la tendencia, variabilidad y estacionalidad de los datos a lo largo del tiempo. Hoy en día el mundo digital nos proporciona no solo una serie de datos sino múltiples series de datos, en nuestro ejemplo tendríamos para cada marca una serie de datos mes a mes.

Analizar estas múltiples series de datos exige aplicar herramientas más complejas que permitan examinar todas las series en su conjunto y de forma simultánea. Una posible herramienta es el análisis de factores dinámicos (DFA), cuyo principal objetivo es ayudar a los investigadores que se enfrentan a datos masivos de corte longitudinal a resumir esa vasta cantidad de información en unos pocos factores de tendencia. Respecto al análisis de series temporales clásico, en lugar de analizar cada serie de manera separada y sacar conclusiones una a una o sobre un conjunto de ellas seleccionándolas de manera arbitraria, DFA ofrece una aproximación estadística robusta para estudiar simultáneamente amplios paneles de series temporales, permitiendo al científico extraer un pequeño número de factores latentes de tendencia comunes del conjunto entero de datos.

Resultados

Aplicamos de nuevo el software KATE [6] para este ejemplo haciendo uso del módulo Dynamic Factor Analysis.

Una vez ejecutada la herramienta y seleccionada una solución de 7 tendencias comunes, al ser la que presenta un valor de ajuste predictivo más adecuado de acuerdo al criterio estadístico BIC, la figura 4.16 muestra el tipo de resultados que nos reportaría el software.

Para cada marca se obtiene un R^2 que indicaría lo bien que estaría representada la marca en la solución global. Hay marcas con un valor muy elevado como por ejemplo Chevrolet (R^2: 0,964) y otras con un valor algo más bajo como Volvo (R^2: 0,491). Por lo general, todas las marcas estarían bien representadas en la solución obtenida.

UNIVERSIDAD DE GRANADA
abiertaugr la formación abierta de la UGR
CC BY NC ND

La interpretación se haría de acuerdo a los coeficientes obtenidos para cada marca en cada tendencia común, identificando los coeficientes con un valor más alto (ya sea en valor + o -) (figura 4.16), y también analizando la evolución del interés de esa tendencia a lo largo del tiempo (figura 4.17):

- La tendencia 1 representaría muy bien el comportamiento de las marcas Audi, Chevrolet, Seat y Volkswagen). Se puede observar que las búsquedas de estas marcas han ido creciendo desde el comienzo del periodo, produciéndose un pico en septiembre de 2015, fecha en la que se descubrió el escándalo de las emisiones de los vehículos del grupo Volkswagen (Dieselgate). Con posterioridad, el interés en las marcas fue disminuyendo ligeramente, aunque se mantuvieron más o menos estables.
- La tendencia 2 recogería la evolución del interés de las marcas Alfa Romeo, Citroen, Fiat, Lancia, Mazda, Nissan, Opel y Renault. Lo relevante es un cierto comportamiento cíclico en el que los primeros meses del año se producen mayor número de búsquedas y en los últimos meses del año las búsquedas sufren un descenso.
- La tendencia 3 es exclusiva de la marca Volvo. Se observan subidas y caídas del interés por esta marca a lo largo del periodo, registrándose también un fuerte pico en 2015 coincidiendo con el escándalo de las emisiones del diésel.
- La tendencia 4 incluye a la marca Toyota. En este caso, el análisis gráfico de la tendencia muestra que ha venido perdiendo volumen de búsquedas en Google, por tanto interés, desde el principio del periodo, hasta el 2016 cuando se produce un fuerte repunte del interés motivado por la apuesta comercial que empezó a realizar la marca por los coches híbridos y eléctricos.
- La tendencia 5 coincide con la marca Dacia. Ha ido aumentando el interés de la población a nivel mundial de manera progresiva, si bien a partir de 2016 se ha producido un pequeño descenso. No obstante, podemos concluir que es una marca que ha ganado atención en el mercado.
- La tendencia 6 incluiría a las marcas Honda, Hyundai, Kia, Mitsubishi y Peugeot. Se trata de marcas de compañías asiáticas, a excepción de Peugeot. Su tendencia a lo largo de los años es claramente descendiente, con altibajos a lo largo del periodo. Podemos concluir que son marcas que claramente han perdido interés entre los consumidores, al menos en términos de búsquedas de información en la Red.
- La tendencia 7 incluye a las marcas Mercedes y Skoda. Presentan una tendencia claramente descendente desde el inicio del periodo. A diferencia de las marcas de la tendencia 6, en este caso apenas hay altibajos y la pérdida de interés parece clara.

UNIVERSIDAD DE GRANADA
abiertaugr la formación abierta de la UGR
CC BY NC ND

	A	B	C	D	E	F	G	H	I
1	Dynamic Factor Analysis								
2		Iterations =	380		BIC =	22521,37		% Change in BIC=	0,00
3								Factor Loadings	
4		R-square	Latent Trend 1	Latent Trend 2	Latent Trend 3	Latent Trend 4	Latent Trend 5	Latent Trend 6	Latent Trend 7
5	ALFA ROMEO	0,515	0,327	**0,667**	0,016	0,573	-0,049	-0,515	-0,117
6	AUDI	0,638	**0,980**	0,361	0,654	0,111	-0,350	0,218	-0,122
7	CHEVROLET	0,964	**-0,381**	0,283	0,183	0,143	0,044	0,179	0,048
8	CITROEN	0,687	0,546	**1,021**	0,168	0,533	0,418	-0,830	0,484
9	DACIA	0,642	0,652	0,836	-0,022	0,550	**0,892**	-0,246	-0,008
10	FIAT	0,671	0,759	**0,877**	0,055	0,525	0,185	-0,372	-0,200
11	HONDA	0,600	0,297	0,605	0,276	0,057	0,011	**-0,744**	0,059
12	HYUNDAI	0,547	0,359	0,311	0,219	0,137	0,177	**-0,480**	-0,282
13	KIA	0,791	0,301	0,414	0,107	0,124	0,201	**-0,530**	-0,410
14	LANCIA	0,622	-0,086	**0,417**	-0,004	0,255	0,035	0,318	-0,043
15	MAZDA	0,873	0,424	**0,445**	0,122	0,033	-0,134	0,035	-0,801
16	MERCEDES	0,907	0,164	0,304	0,046	-0,052	0,111	-0,010	**-0,904**
17	MITSUBISHI	0,698	-0,254	0,165	-0,012	0,148	0,267	**-1,801**	1,070
18	NISSAN	0,575	0,609	**0,637**	0,102	0,322	0,058	-0,171	-0,434
19	OPEL	0,720	0,773	**1,081**	0,272	0,676	0,339	-0,594	0,063
20	PEUGEOT	0,709	0,172	0,617	0,242	0,403	0,005	**-0,901**	-0,205
21	RENAULT	0,733	0,857	**0,877**	0,097	0,447	0,234	-0,762	0,332
22	SEAT	0,596	**1,343**	0,700	0,537	0,494	0,342	0,044	0,356
23	SKODA	0,742	0,147	0,480	0,221	0,369	0,054	-0,470	**-0,763**
24	TOYOTA	0,962	0,515	0,483	0,113	**1,211**	0,168	-0,264	-0,787
25	VOKSWAGEN	0,989	**1,523**	0,186	1,129	-0,102	0,019	0,038	0,127
26	VOLVO	0,491	0,504	0,208	**0,548**	-0,118	0,048	-0,128	-0,331

Figura 16. *Resultados de la aplicación de la herramienta. Fuente elaboración propia.*

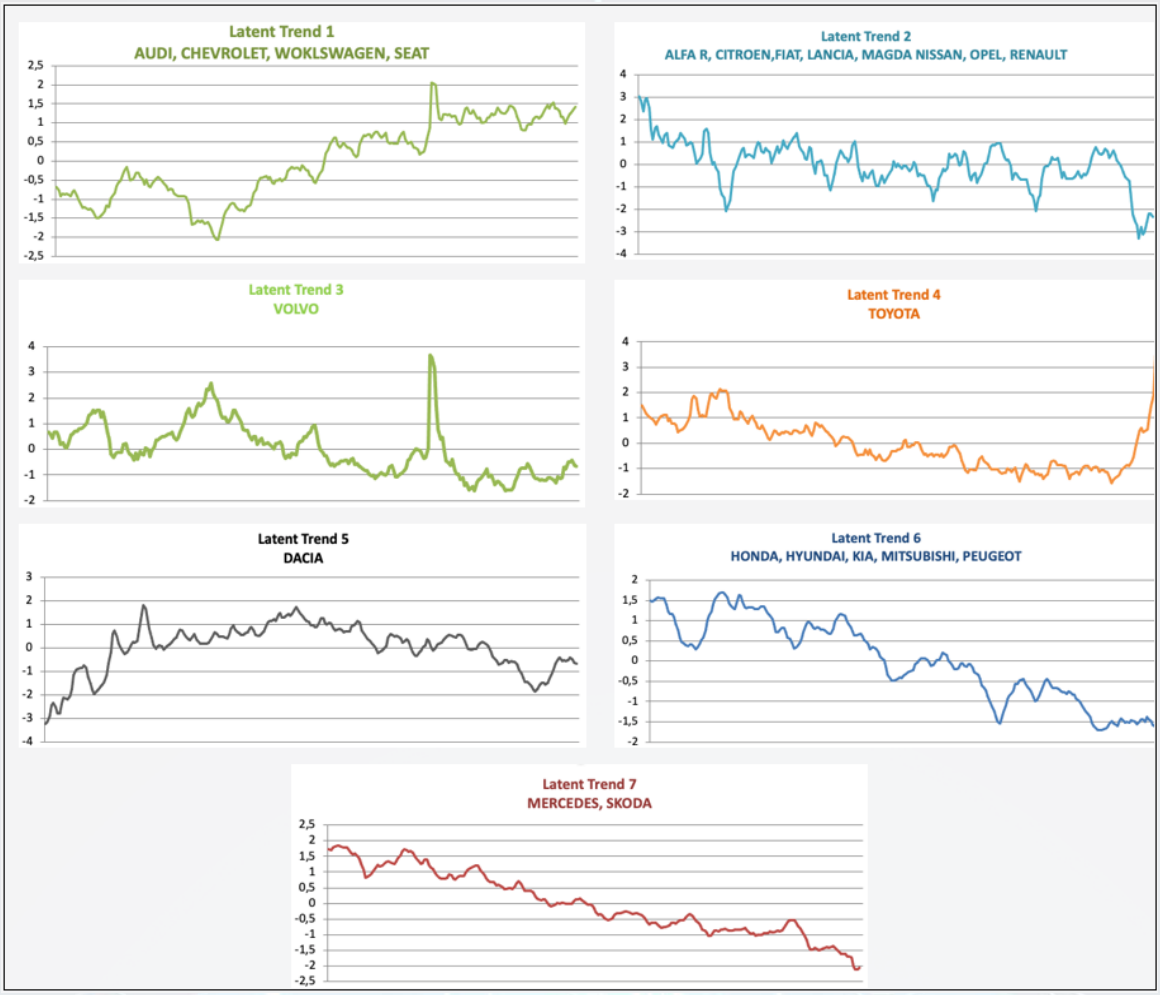

Figura 17. *Tendencias comunes. Fuente elaboración propia.*

UNIVERSIDAD DE GRANADA
abiertaugr
la formación abierta de la UGR
CC BY NC ND

A modo de conclusión, esta herramienta DFA permite examinar grandes volúmenes de datos de corte longitudinal y resumir toda esa información en unas cuantas tendencias que explicarían el comportamiento de las marcas examinadas. Por tanto, las pepitas de información (insights) que proporciona resultan de gran utilidad para la toma de decisiones en negocios y marketing, al poder monitorizar la evolución del interés de una marca a lo largo del tiempo.

4. Conclusiones

Tras la lectura de esta unidad el lector habrá sido capaz de entender en qué consiste la ciencia de datos y el valor que aporta para las empresas, así como cuáles son las herramientas que existen. También se ha analizado la relación y diferencias entre la ciencia de datos y la IA y cómo ambas sirven para poder realizar análisis de datos para extraer valor de los mismos.

Los ejemplos prácticos analizados no persiguen explicar con detenimiento las herramientas de ciencia de datos aplicadas, sino más bien de que el lector sea consciente de cómo pueden ayudar a la generación de valor a las empresas y el tipo de resultados que pueden aportar.

5. Bibliografía

[1] https://www.computerweekly.com/es/definicion/Inteligencia-artificial-o-IA

[2] https://data-flair.training/blogs/data-science-vs-artificial-intelligence/

[3] https://openwebinars.net/blog/que-es-el-machine-learning-y-como-funciona/

[4] https://data-flair.training/blogs/data-science-use-cases/

[5] https://www.master-data-scientist.com/en-que-consiste-el-text-mining/

[6] Kamakura's Analytic Tools for Excel: https://www.katexcel.com

[7] https://trends.google.es/trends

UNIVERSIDAD DE GRANADA
abiertaugr la formación abierta de la UGR
CC BY NC ND