

## CAPÍTULO I

## Importancia del *big data* en *marketing* digital

Daniel Peña\*

Las empresas que han enfocado su actividad comercial a través de Internet recogen de forma automática información sobre sus clientes que puede utilizarse para mejorar sus estrategias de *marketing* digital. Este trabajo describe cómo los nuevos datos masivos, el *big data*, que incluye además de los datos tradicionales los basados en imágenes y sonidos digitales, está transformando los procesos de venta *online*. Estos cambios se ilustran con tres herramientas comerciales desarrolladas con datos masivos. La primera, los sistemas de recomendación, es nueva y constituye un buen ejemplo de las oportunidades que ofrece el análisis de la información sobre compras previas de los clientes. La segunda, el seguimiento de la lealtad mediante el cálculo de la probabilidad de abandono de los clientes, tiene una larga tradición, pero se ha perfeccionado y redefinido en el nuevo contexto de datos masivos. La tercera, el análisis de las características de las redes de clientes, se está introduciendo en la actualidad y tiene un amplio potencial de desarrollo. Permite describir las relaciones entre clientes y entender mejor la estructura de los usuarios, construir nuevas variables de alto valor predictivo y diseñar nuevas estrategias de comercialización en *marketing* digital. Su potencial se ilustra con un ejemplo de los resultados al analizar la red de clientes de un importante banco. El artículo concluye con una revisión de otras herramientas de *marketing* digital que van a transformarse por el uso de los datos masivos.

*Palabras clave:* nuevos datos, sistemas de recomendación, clasificación de clientes, lealtad de clientes, probabilidades de abandono, redes de clientes.

---

\* Con el apoyo parcial de la Agencia Nacional de Evaluación de la Calidad y Acreditación con referencia PID2019-109196GB-I00.

## 1. INTRODUCCIÓN

El nombre *marketing* digital comenzó asociándose a la comercialización de productos y servicios utilizando canales digitales, pero ha evolucionado para incluir todos los aspectos ligados a la venta, incorporando la colaboración con otras empresas, clientes y organizaciones para crear valor para el accionista (Kannan, 2017). Su desarrollo es relativamente reciente, ya que hasta 1990 Internet no podía utilizarse para fines comerciales y el uso de la web estaba restringida a usos científicos y militares. En 1993 la web fue adaptada para ser aprovechada por los usuarios privados para la venta y una de las primeras compañías basadas en Internet fue Amazon en 1994. Desde entonces, la importancia de Internet en los negocios ha crecido de forma espectacular. Las mayores empresas del mundo, y de mayor crecimiento, han dejado de ser las compañías petroleras o de automoción, y han sido sustituidas por las organizaciones que, a través de la web, obtienen información constante sobre sus clientes, como Google, Amazon o Facebook.

Una de las consecuencias de Internet y de los avances en tecnologías de la comunicación es la aparición de los datos masivos, los llamados *big data*. Véase Marx (2013) para su impacto general y Erevelles, Fukawa y Swayne (2016) para la transformación del *marketing*. Los nuevos datos masivos han dado lugar a mejoras en los procesos comerciales, como los sistemas de recomendación y de retención de clientes, y han surgido nuevos modelos de negocio basados en ellos, como las grandes compañías de transporte de viajeros que no tienen un solo vehículo ni medio de locomoción. Por otro lado, los mismos datos se han convertido en un producto valioso, que, en bruto o una vez procesados para convertirlos en información, pueden venderse a posibles usuarios.

Muchas organizaciones recogen hoy en día información sobre sus clientes. Los datos se generan tanto en los procesos de venta como en el resto de las interacciones entre las empresas y los usuarios de sus productos y/o servicios. También existen en Internet otros datos sobre los clientes, que provienen del uso de las redes sociales, los teléfonos móviles, sensores de varios tipos etc., donde los usuarios han ido dejando un rastro de sus actividades. Las empresas que ofrecen productos *online* y que utilizan Internet para la comunicación con sus clientes, como las empresas de servicios financieros, de seguros, de telecomunicación y ocio y, en general, de venta de productos y servicios por la red (servicios de transporte, de alojamiento y turismo etc.), pueden utilizar estos datos para mejorar la venta de sus productos y servicios, mejorando con ello su posición en el mercado (véase, por ejemplo, Ryan, 2016; Sayyad *et al.*, 2018 y Saura, 2021).

Los datos brutos disponibles, como tiempo de visita de una página web, mensajes escritos en una red social o geolocalización de una venta, pueden convertirse en activos de valor comercial sobre las preferencias y deseos de los clientes con métodos de Ciencia de los Datos, o *Data Science*. Este enfoque incluye procesos de aprendizaje automático (*machine learning*, *ML* por sus siglas en inglés) e inteligencia artificial (o *AI*, *artificial intelligence*), combinados con métodos estadísticos (Galeano y Peña, 2019a). Los datos utilizados provienen no solo de mediciones numéricas y de tablas de variables, sino, también, de imágenes, textos o

señales de audio digitalizadas. La inteligencia artificial se ha concentrado en enseñar a máquinas digitales a aprender con los sentidos humanos, como la vista (reconocimiento de imágenes y clasificación de textos) y el oído (análisis de señales) y este conocimiento se incorpora a las mediciones tradicionales para realizar predicciones o clasificaciones de los clientes.

La información obtenida de estos datos masivos nos permite: a) prever el comportamiento futuro de un cliente en función de sus datos pasados; b) clasificar clientes en grupos homogéneos por sus comportamientos y utilizar esta clasificación para personalizar las ofertas y mejorar las predicciones individuales; y c) entender las variables clave que influyen en el comportamiento de los clientes para diseñar nuevas estrategias y canales de venta. Estos nuevos datos pueden no solo ayudarnos a mejorar lo que hacemos sino, sobre todo, a diseñar nuevos productos y estrategias comerciales para: 1) atraer, retener y satisfacer a los clientes; 2) utilizar el *marketing mix* para maximizar los resultados de las inversiones; 3) personalizar las estrategias de *marketing* para cada cliente en función de sus necesidades, y 4) aumentar la seguridad y privacidad de la venta *online* (Wedel y Kannan, 2016).

Este trabajo comienza describiendo cómo las imágenes y audios se convierten en ficheros digitales para ser transformados en información utilizable en el *marketing* digital. A continuación, se analizan tres herramientas comerciales de contrastada eficacia basadas en aprovechar la información sobre los clientes. Su estructura es la siguiente. En la sección 2 describimos brevemente cómo las imágenes y vídeos, así como las señales de audio, se convierten en datos digitales utilizables para la clasificación y la predicción. En la sección 3 analizamos una de las herramientas de *marketing* digital más conocidas y utilizadas: los sistemas de recomendación. La sección 4 presenta las nuevas técnicas de retención de clientes basadas en el cálculo de probabilidades de abandono y la toma de decisiones para garantizar su continuidad. La sección 5 está dedicada a una nueva herramienta para el *marketing* digital: al análisis de los clientes en red. Las redes de clientes permiten medir mejor la importancia relativa de cada uno, entender las relaciones entre ellos, desarrollar nuevas estrategias comerciales y predecir mejor variables clave para la rentabilidad del negocio. Finalmente, la sección 6 incluye algunas conclusiones finales.

## 2. LOS NUEVOS DATOS EN *MARKETING* DIGITAL: IMÁGENES Y SEÑALES

La vista y el oído son dos de las fuentes principales de recogida de información del ser humano y el *marketing* digital ha incorporado los datos conseguidos con estos sentidos como posibles herramientas comerciales. Es previsible que los otros tres sentidos adicionales, olfato, tacto y gusto, tengan también un impacto creciente en el *marketing* digital del futuro y el llamado *marketing* sensorial (Hultén, 2017; Hussain, 2019) se vislumbra como una nueva oportunidad en muchas áreas.

Las imágenes han sido una herramienta fundamental de venta para mostrar un producto de forma atractiva para el consumidor y desde el comienzo de la fotografía en el

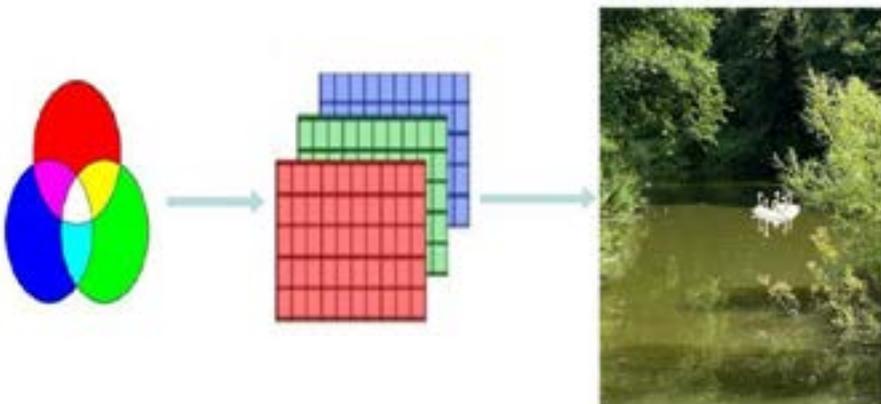
siglo XIX se han utilizado en publicidad. Si una imagen vale más que mil palabras, una adecuada presentación visual de un producto o servicio puede ser mucho más efectiva que una detallada descripción de sus propiedades o características. Por esta razón, durante el pasado siglo la fotografía ha sido una importante herramienta comercial y su uso generalizado para estimular las ventas ha contribuido a su gran desarrollo. En la actualidad, una adecuada imagen del producto es decisiva en los portales de compra y tiendas virtuales a las que se accede por Internet.

El desarrollo de las técnicas de análisis de imágenes digitales a finales del siglo XX ha abierto otras oportunidades para el uso de las fotografías y los vídeos en la promoción y venta de productos y servicios. Por ejemplo, Velasco *et al.* (2015) encontraron que una correspondencia entre el color del envase y el producto contenido estimula la venta en supermercados: cuando el color del paquete era congruente con el producto que contiene (rojo/tomate) el cliente lo encontraba más fácilmente, potenciando su venta, mientras que esta ventaja no aparecía sin esta congruencia (amarillo/tomate). Este resultado ilustra que medir la percepción que produce una imagen en un posible cliente puede ayudar a mejorar su etiquetado y su diseño.

Las imágenes hoy son una fuente de información frecuente sobre el consumidor y se están utilizando para desarrollar nuevas estrategias de venta. Fotos de los productos comprados o de las páginas visitadas por un cliente o grabación por vídeo de su comportamiento en el punto de venta, proporcionan datos valiosos sobre las personas que eligen un producto o servicio. Su análisis puede descubrir patrones de comportamiento de ciertos tipos de cliente, utilizables luego comercialmente. Otro uso de importancia creciente de las imágenes es para identificar y localizar productos: subiendo a un buscador o plataforma de compra una foto

Figura 1.

### Construcción de una imagen digital con el sistema RGB



Fuente: Elaboración propia.

de un producto, por ejemplo de un vestido, el buscador puede automáticamente identificar el producto y ofrecer su venta. Para ello, se requiere la capacidad de reconocer y clasificar imágenes mediante técnicas de inteligencia artificial.

Una imagen digital es un conjunto de muchos números dispuestos en filas y columnas en una estructura rectangular que llamamos matriz. Cada uno de estos números indica el color de una diminuta superficie cuadrada de la imagen que llamamos píxel. Estas imágenes digitales se denominan de mapa de bits, o *bitmap* y su resolución se define por el número de píxeles que contienen, bien en densidad, píxeles por pulgada (ppi) o por centímetro (pcm), o bien en número de píxeles totales. Por ejemplo, una resolución de un megapíxel equivale a un millón de píxeles, o a un cuadrado con 1.000 píxeles en cada lado.

Una imagen de un tamaño dado será tanto más nítida cuantos más píxeles contenga, y si aumentamos su tamaño, manteniendo el mismo número de píxeles, comenzará a verse borrosa. El ojo humano, como el de otros seres vivos, tiene un límite en la cantidad de píxeles que puede apreciar, aunque este límite es muy alto con relación a las cámaras actuales. El ojo no toma imágenes fijas, como una cámara, sino que está en continuo movimiento. Se estima que el ojo humano promedio tiene alrededor de 576 megapíxeles de resolución.

Cada píxel tiene un color uniforme y la sensación de color en la imagen se forma como resultado de integrar visualmente, en la retina, las variaciones de color y luminosidad entre píxeles vecinos. El color de cada píxel puede definirse de dos formas principales. En las figuras en blanco y negro, o en tonos de grises, el color se define por una matriz que contiene un número asociado para cada píxel, que es la intensidad de gris. En las figuras en color el método más utilizado, llamado RGB, es utilizar la intensidad de tres colores básicos, rojo (*red*), verde (*green*), y azul (*blue*) para generar mediante la combinación de estos tres colores cualquier otro. Para ello, se utilizan tres matrices del mismo tamaño y el número que contienen en cada casilla representa la intensidad de uno de los tres colores R, G y B. La combinación de estos tres números define el color asociado a cada píxel de la imagen. La figura 1 ilustra esta representación.

Para clasificar imágenes se utilizan sus píxeles, que pueden ser muchos, y normalmente los archivos *bitmap* se comprimen para procesarlos. Entre los sistemas de compresión más utilizados están el formato jpg, que desecha primero información no visible y el formato png que crea archivos algo mayores que el jpg.

## 2.1. Clasificación de imágenes

La clasificación de imágenes digitales consiste en asignar cada imagen a una clase. Estas clases pueden estar definidas *a priori*, por ejemplo, hombre o mujer, y entonces hablamos de clasificación supervisada o de discriminación. En estos casos las imágenes disponibles para construir el clasificador han sido previamente asignadas a los grupos y llevan cada una la etiqueta del grupo al que pertenecen. El objetivo es clasificar en uno de los grupos estableci

dos una nueva imagen que llega sin etiqueta. Por ejemplo, tenemos muchos datos de compras previas de personas que hemos identificado como hombres o mujeres y dado un conjunto de compras sin etiqueta queremos clasificar al comprador como un hombre o una mujer. Un problema diferente aparece cuando los datos disponibles no tienen etiquetas, de manera que los posibles grupos no están definidos *a priori*. El objetivo es ahora encontrar los grupos, si existen, a partir de los datos y hablamos entonces de clasificación no supervisada o análisis clúster. Una tercera situación, mezcla de las dos anteriores, es cuando tenemos datos híbridos, que es el caso de la clasificación semisupervisada. Algunos de los datos disponibles para construir el clasificador llevan etiquetas y otros carecen de ellas. Hay, por lo tanto, ya unos grupos definidos, como en la clasificación supervisada, pero nos planteamos la posibilidad de ampliar su número considerando nuevos grupos analizando los datos, como en el análisis clúster.

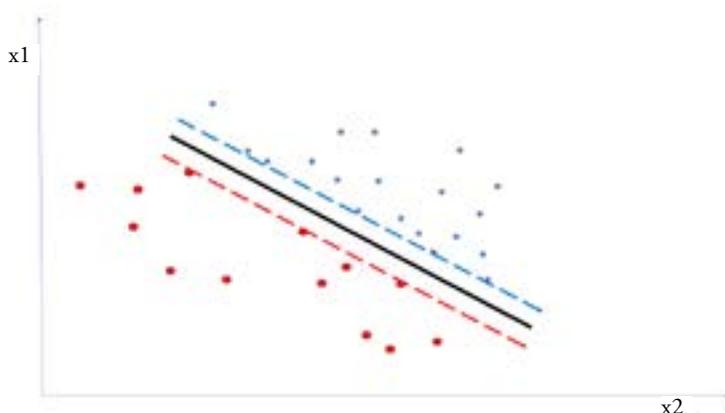
Consideremos inicialmente el problema de clasificación supervisada, o análisis discriminante, entre un grupo de clases previamente definido. Por ejemplo, clasificar una fotografía como incluyendo a personas o no. Para ello, los valores de los píxeles de la imagen se colocan formando un vector de tantos componentes como píxeles, de manera que el vector contiene en orden los valores de las filas (o de las columnas). Otra alternativa es resumir este vector en un conjunto de valores numéricos representativos de la imagen, como la media de los valores de los píxeles, la varianza o la correlación espacial entre los píxeles, y formar un vector de características asociado a cada imagen  $x$ , que tendrá habitualmente una dimensión mucho menor que el número total de píxeles en la imagen. Dada una nueva imagen calculamos su vector de características y lo comparamos con los de las imágenes en los grupos existentes.

La comparación puede hacerse de dos maneras: de forma global, comparando las características de la imagen con las medias de las características de cada grupo, o de forma local, comparando la imagen con las más similares de cada uno de los grupos. En ambos casos, la clasificación final se realiza calculando una medida de distancia entre el vector de características de la imagen a clasificar y las imágenes incluidas en cada uno de los grupos, asignando la imagen al grupo más próximo con un cierto criterio. Además de realizar la clasificación, podemos calcular una probabilidad de que la imagen pertenezca al grupo fijado, lo que nos define la incertidumbre en la clasificación.

Para concretar, supongamos el caso más simple de dos grupos. Disponemos de una muestra de entrenamiento con  $n_1$  elementos del primer grupo y  $n_2$  del segundo. Cada imagen se representa por un vector de dimensión  $p$  que contiene los valores de los píxeles de la imagen. Llamaremos  $\mathbf{x}_1, \dots, \mathbf{x}_{n_1}$  a los vectores del primer grupo y  $\mathbf{y}_1, \dots, \mathbf{y}_{n_2}$  a los del segundo. Tenemos una nueva imagen  $\mathbf{z}$  que queremos clasificar. Los métodos globales, que son los más utilizados tradicionalmente, calculan una medida de distancia entre los valores de la nueva imagen y las medias de los grupos, pero este procedimiento puede no ser muy útil para clasificar imágenes. Por ejemplo, la media de los píxeles puede ser un valor poco relevante para entender la separación entre los grupos y sería mucho más efectivo comparar los valores de un conjunto de píxeles determinados de ambos grupos. Por ejemplo, supongamos que las imágenes de la muestra no presentan separación entre los píxeles salvo para las dos primeras coordenadas,  $x_1, x_2$ , donde se observa la representación de la figura 2. Un análisis basado en todos los

Figura 2.

### Línea de separación de dos grupos de datos con máquinas de vector soporte (SVM)



Fuente: Elaboración propia.

píxeles y su valor medio, u otro resumen del conjunto de píxeles, no será efectivo para separar las imágenes, pero una combinación lineal de las dos variables representadas en la figura 2 permite separar ambas poblaciones. En este caso será más efectiva una separación local. Una técnica que utiliza este enfoque para realizar esta separación es una máquina de vector soporte o *support vector machine* (SVM). Este procedimiento trata de buscar el plano frontera entre los dos conjuntos de puntos que maximiza la separación entre ambos grupos. En la figura 2 el plano es una recta y representa la separación óptima que puede encontrarse entre ambos conjuntos de puntos.

Un tercer procedimiento es transformar el problema de la clasificación en un problema de predicción de la clase. Por ejemplo, con dos clases podemos representar la primera con la variable respuesta 0 y la segunda con la variable respuesta 1 y el problema es, conocidos los valores de las características, prever la clase. Definimos una variable a predecir,  $y$ , que puede tomar los valores cero y uno. Para predecir esta variable disponemos de una muestra de entrenamiento de imágenes con datos del valor de la respuesta  $y$  en cada imagen y de un vector de variables  $x$  que representa los píxeles de cada imagen. Los modelos de predicción tradicionales, como la regresión logística, suponen que, en el caso particular de dos clases, la probabilidad de pertenecer a la clase  $y = 1$  viene dada por una función del tipo:

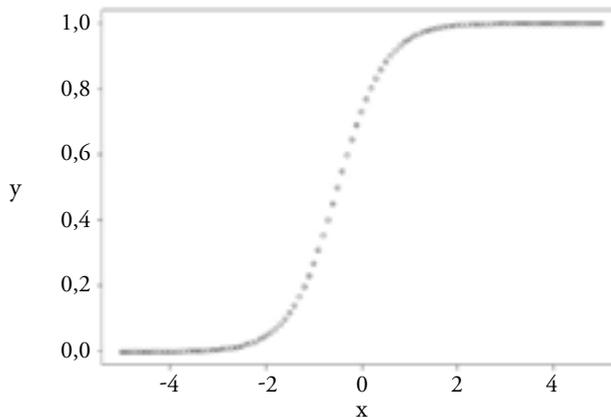
$$p(y = 1) = g(\mathbf{x}'\mathbf{b}) = g(b_0 + x_1b_1 + \dots + x_pb_p) \quad [1]$$

donde  $(x_1, \dots, x_p)$  son las variables explicativas, los píxeles de la imagen, y los coeficientes  $(b_0, b_1, \dots, b_p)$  tienen que estimarse con los datos. La probabilidad de la otra clase se obtiene con  $p(y = 0) = 1 - p(y = 1)$ , y  $g$  es una función no lineal que suele tomarse como la función logística, que aparece representada para el caso de una única variable explicativa en la figura 3. El gráfico muestra que la probabilidad de pertenecer a la clase  $y = 1$  aumenta con los valores

de  $x$ , primero lentamente, casi linealmente en un intervalo, y esta probabilidad se estabiliza a partir de valores altos de  $x$ . Con muchas variables el efecto reflejado en la figura 3, puede interpretarse como el de cada variable cuando el resto de las otras variables toma valores fijos. La pendiente en la zona de crecimiento depende del coeficiente de la variable,  $b_i$ , y será en general distinto para las diferentes variables.

Figura 3.

### Modelo logístico entre dos variables



Fuente: Elaboración propia.

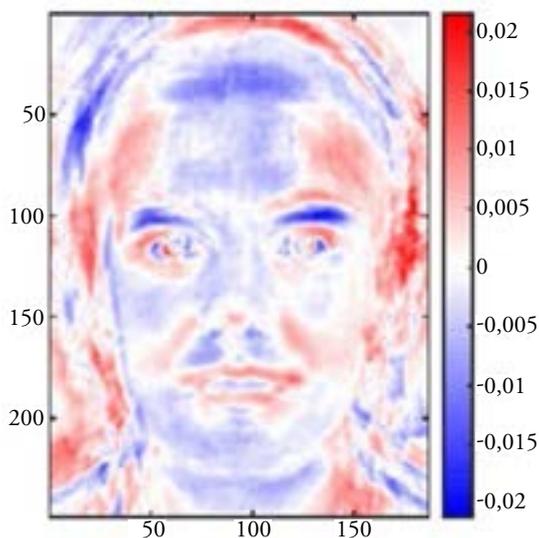
Este modelo puede generalizarse de forma que la relación entre la variable respuesta y la variable  $x$  sea más compleja utilizando redes neuronales. Una red neural supone muchas capas de modelos similares al [3], de manera que el efecto total se representa como superposición de muchas respuestas de tipo [3]. De esta manera puede demostrarse que es posible aproximar cualquier relación compleja entre la variable respuesta,  $y$ , y las explicativas  $x$ . Los métodos de clasificación mediante redes neuronales con varias capas se conocen como métodos de aprendizaje profundo. Véase, por ejemplo, Levi y Hassner (2015) para la clasificación de género y edad con redes neuronales.

En la sección siguiente presentamos un ejemplo simple de clasificación del género de una persona a partir de una imagen fotográfica.

#### 2.1.1. Un ejemplo de clasificación por sexo

La clasificación de un rostro humano correspondiente a un hombre o una mujer se utiliza frecuentemente en *marketing* en el análisis de vídeos que representan transacciones comerciales en puntos de venta. Este proceso requiere primero la identificación de los píxeles correspondientes al rostro en la imagen digital y, después, la clasificación de la cara como hombre o como mujer. Nos centraremos solo en este segundo aspecto, que ha sido muy estu-

Figura 4.

**Importancia de los píxeles para la discriminación entre hombres y mujeres**

*Fuente:* Elaboración propia.

diado desde el inicio de la estadística. Uno de los primeros trabajos en esta línea fue debido a Galton (1910), el inventor de la idea de regresión, que utilizó variables como ángulos y distancias entre las características de los rasgos faciales, (nariz, boca, etc.) para clasificar nuevas caras. Galton obtuvo para cada rostro un conjunto de medidas que comparaba con una base de datos de estas medias en rostros de hombres y mujeres. Desde entonces, este problema ha sido muy estudiado en distintos contextos, véase por ejemplo Zhao *et al.* (2003) y Kawulok, Celebi y Smolka (2016). Un problema importante cuando se trabaja con muchas imágenes es reducir la dimensión de cada matriz de píxeles sin perder información relevante para la clasificación. Existen procedimientos generales para comprimir imágenes, pero el proceso puede hacerse de manera más efectiva si tenemos en cuenta el tipo de imágenes que queremos comprimir y las características a mantener, vease Benito y Peña (2005) para la reducción de la dimensión en imágenes digitales de rostros.

Benito *et al.* (2017) han comparado distintos métodos de clasificación para rostros humanos y determinado las zonas de la cara humana más importantes para esta clasificación. La figura 4, tomada de Benito *et al.* (2017), muestra estas zonas: las zonas de mayor color son las de mayor poder discriminante y las zonas blancas las que tienen poca utilidad para la discriminación. El color azul indica zonas importantes en el rostro femenino y el rojo en el masculino. Las zonas más interesantes para la discriminación son las cejas, los ojos, la nariz, los labios, el mentón y la forma de la cabeza, donde la presencia de pelo corto o largo tiene un importante papel.

Shan (2012) presenta otras estructuras para la clasificación de género y Khan *et al.* (2020) aplican redes neuronales a este problema.

## 2.2. Análisis de señales de audio

La voz es una fuente importante de información en *marketing* digital. Por ejemplo, la atención telefónica actual se realiza habitualmente utilizando un programa de reconocimiento de voz, que clasifica el tipo de llamada para proporcionar la información solicitada. Si la información requerida es simple, se activa una respuesta codificada, en otro caso, la llamada se transfiere a un operador humano. La clasificación de las llamadas por su contenido requiere su análisis digital para comprender la voz humana. Para ello, el sonido captado por un micrófono, que recoge las vibraciones sonoras a través de una membrana, se convierte en señales eléctricas que son enviadas a un ordenador que las digitaliza, es decir, convierte estas señales en datos digitales como secuencias de ceros y unos, a través de la placa de sonido. El resultado obtenido puede representarse como una serie temporal o una función a lo largo del tiempo. Estas secuencias digitales se convierten de nuevo en sonido mediante altavoces, como indica la figura 5.

Figura 5.

### Representación digital del sonido



Fuente: Elaboración propia.

El reconocimiento de voz requiere definir los fonemas o sonidos básicos de un idioma, generalmente asociados a las letras que lo forman y a las combinaciones de letras, e identificarlos en la secuencia para formar palabras. Para reconocer cada fonema se utiliza, en primer lugar, su forma digital habitual. Cada fonema se transforma en una secuencia específica de ceros y unos en su representación digital que caracteriza su sonido y se busca esta secuencia en un archivo de sonidos digitales. Este es un problema de clasificación similar al estudiado para clasificar los rostros humanos, pero ahora en lugar de dos clases, hombre y mujer, tenemos que clasificar cada sonido en uno de los numerosos fonemas del idioma.

La mayoría de los idiomas tienen entre 20 y 35 fonemas. Por ejemplo, el español tiene 24 fonemas (5 vocales y 19 consonantes) y es uno de los idiomas latinos más simples: por ejemplo, el italiano tiene 34 fonemas y el francés, 37. Los idiomas del norte de Europa son más complejos que los derivados del latín: el inglés tiene 44 fonemas y el danés llega hasta 50 (30 vocales y 20 consonantes), y existen idiomas en África con 85 fonemas. Un problema adicional es que es frecuente que en el idioma existan distintos acentos, o forma de pronun-

ciación, por lo que distintas personas lo emiten de forma diferente. Por tanto, hay que definir la clase de sonidos que corresponden a cada fonema, que a veces es muy amplio, y que llamaremos la clase de sonidos del fonema. El objetivo es, por tanto, clasificar las señales digitales en estas clases de fonemas.

La clasificación de sonidos se realiza de forma más efectiva con métodos locales, utilizando una amplia base de datos de pronunciaciones del idioma. Estos diccionarios de palabras (*speech database*), se construyen grabando previamente a muchos locutores de un idioma. El sistema clasifica los conjuntos de fonemas dentro de ese diccionario para identificar, o escuchar, las palabras pronunciadas. Estos métodos mejoran además con la experiencia de su uso, porque se puede ir ampliando la base de datos de sonidos al aplicar el clasificador cuando es posible comprobar si esta ha sido correcta y se ha clasificado con éxito. Por esta razón, es frecuente que la máquina pida la confirmación al cliente sobre su comprensión, de manera que puede etiquetar los sonidos escuchados en la clase correcta e incorporarlos a la base de datos, mejorando su funcionamiento con la experiencia. En este sentido el sistema aprende con su uso, propiedad muy deseable en un clasificador. Por ejemplo, si utilizamos una metodología de clasificación como las máquinas de vector soporte que se describieron en la sección anterior, la base de elementos con etiqueta bien clasificados aumenta continuamente mejorando la precisión del clasificador.

La clasificación inicial de los sonidos, cuando la base de datos disponible no es muy grande, puede mejorarse utilizando, además, las probabilidades de transición entre fonemas del idioma, es decir, la frecuencia con la que cada fonema aparece a continuación de uno dado. El reconocimiento de voz puede hacerse más preciso cuando es posible añadir un sistema de vídeo que analiza los labios del emisor para identificar los fonemas que utiliza. La evidencia demuestra que mezclando ambos métodos, el visual y el auditivo, se obtiene un sistema más potente. Véase Gong (1995), Padmanabhan y Premkumar (2015) y Nassif *et al.* (2019) para una descripción de los procedimientos utilizados en el reconocimiento de voz.

Los sistemas de reconocimiento de voz han dado lugar a los asistentes digitales, que, además de contestar preguntas de los usuarios, como Alexa de Amazon o Siri de Apple, pueden realizar acciones de asistencia como poner música, encender las luces o buscar el significado de una palabra en un diccionario. También comienzan a incorporarse en los sistemas de recomendación, que describimos en la sección siguiente.

### 3. SISTEMAS DE RECOMENDACIÓN

La selección de productos por un cliente dentro de una amplia oferta de los mismos se simplifica con ayuda de los sistemas de recomendación. Estos sistemas ofrecen al usuario productos que pueden interesarle y le ayudan a encontrar otros artículos atractivos para su perfil entre los disponibles. El rápido crecimiento de Internet y del comercio electrónico han propiciado la proliferación de estos sistemas de *marketing* digital. Véase Smith y Linden (2017) para

un análisis de la experiencia de Amazon, pionera en este campo y Kumar y Thakur (2018) para una revisión amplia de estos sistemas.

Los sistemas de recomendación (SR) pueden clasificarse en tres categorías dependiendo del procedimiento utilizado para construirlos: 1) por contenido; 2) filtros de colaboración, y 3) métodos híbridos. Los SR por contenido se basan en la historia pasada de compras similares del cliente, por ejemplo, recomendar una película histórica a clientes que en el pasado han elegido este tipo de películas. Es el procedimiento más simple y el que utilizaría un vendedor familiarizado con las compras previas de un cliente. Se aplica frecuentemente en las páginas web y en las recomendaciones de noticias. Los SR por filtros de colaboración requieren más información, porque utilizan para hacer la recomendación las compras realizadas por otros clientes con características similares al actual, o de productos similares a los que el cliente ha comprado, y pueden, por tanto, proponer artículos en los que el cliente no ha pensado y que desconoce. Este método trata, por tanto, de anticiparse a los deseos del cliente y de hacerle recomendaciones de productos cuya existencia ignora. Los sistemas híbridos combinan la historia de compras del cliente con las de otros clientes similares para generar las recomendaciones.

Los filtros colaborativos comenzaron a utilizarse en los años noventa cuando los científicos de datos de Amazon propusieron un sistema de recomendación, Linden *et al.* (2003), que ha tenido gran difusión. Los métodos más utilizados inicialmente se basaban en asociar a cada cliente un conjunto de “vecinos” definidos como personas con compras similares a las suyas. A cada vecino de un cliente se le asigna un número que mide su grado de afinidad, que depende de su distancia entre su vector de compras y el del cliente para el que construimos la recomendación, utilizando distintas medidas de proximidad para asociarles. La recomendación se hace considerando el grupo de clientes más próximos al actual y recomendando primero los productos más frecuentemente comprados en ese grupo y que no han sido todavía adquiridos por el cliente.

La idea de Linden, Smith y York (2003) para construir el SR de Amazon fue buscar similitudes, no entre los clientes, sino entre los productos. Partiendo de los datos de compras, se construye una matriz donde aparecen todos los productos y clientes y sus compras. Con esta matriz podemos calcular probabilidades de compra de productos y de pares de productos. Dos productos,  $X$ ,  $Y$ , están relacionados cuando su probabilidad de compra conjunta  $P(X, Y)$  es mucho mayor que el producto de las probabilidades de sus compras individuales  $P(X)P(Y)$ . Con esta idea básica se establece una asociación entre productos. Naturalmente esta idea de asociación requiere ajustes para tener en cuenta los tiempos de compra, disponibilidad etc., pero es la base del método. De esta manera, para cada cliente se crea una base de productos “vecinos” a los ya comprados, que es la base de las recomendaciones.

La ventaja de este enfoque es que el énfasis del procedimiento es fijarse en los productos comprados, no en los compradores, con lo que se capta mejor la esencia del impulso de compra. Por ejemplo, dos personas pueden compartir su afición a la pesca y muchos de los aficionados a la pesca compran automóviles, que utilizan para ir a pescar. Por tanto, es natural que un sistema basado en la asociación pesca/coche recomiende a un pescador

que no conduce artículos para vehículos. Esto no ocurre al utilizar los artículos comprados por la persona objeto de la recomendación. El método tiene además la ventaja de que los cálculos intensivos, las probabilidades de compra y las medidas de asociación, se realizan *offline*, con lo que puede aplicarse de forma rápida en el momento de hacer la sugerencia. La aplicación después a millones de clientes puede hacerse de manera muy rápida, así como la actualización de las recomendaciones con nuevos datos.

Los sistemas de recomendación han sido criticados por el uso de información confidencial. Desde 2006 a 2009, Netflix estableció un premio de 1.000.000 de dólares al equipo que con una base de datos consistente en un conjunto de 100 millones de evaluaciones anónimas de películas pudiese generar recomendaciones que superasen al menos en un 10 % a su sistema de recomendación, basado en filtros colaborativos. El mejor sistema premiado en 2007 utilizaba una combinación de 107 métodos distintos. Sin embargo, ese mismo año investigadores de la Universidad de Texas identificaron a los usuarios supuestamente anónimos que habían dado las evaluaciones de las películas cruzando los datos de Netflix, que efectivamente eran anónimos, con los abiertos del Internet Movie Database de Amazon. Esto llevó a una demanda contra Netflix por difundir información privada y a la cancelación del premio. Desde entonces, la privacidad de los sistemas de recomendación ha sido objeto de polémica, especialmente en el campo de la salud, donde este tema es especialmente sensible, véase por ejemplo Katzenbeisser y Petkovic (2008).

A pesar de estas críticas, el éxito de estos sistemas ha contribuido a su difusión. Amazon estima, Smith y Linden (2017), que un 30 % de sus ventas se generan por los sistemas de recomendación y Netflix ha estimado que el 80 % de las películas que ven sus usuarios son consecuencia de sus sugerencias. Los sistemas híbridos son muy utilizados en la actualidad, combinando métodos de predicción basados en modelos con técnicas de clúster y asociación para clientes y productos. Los sistemas de recomendación se han demostrado además útiles para desarrollar la lealtad del cliente. Por ejemplo, Wang *et al.* (2009) han propuesto un procedimiento de sugerencias para empresas de redes inalámbricas para evitar la deserción de clientes, un problema que estudiaremos en la sección siguiente.

Los sistemas de recomendación están incorporando el uso de voz y combinándose con los asistentes digitales para proporcionar sugerencias y asistencia de forma verbal a solicitudes varias del cliente, como encontrar un restaurante vegetariano, reservar una habitación en un hotel o comprar las entradas de un espectáculo.

#### 4. SEGUIMIENTO DE LA LEALTAD DE LOS CLIENTES

En un entorno de alta competencia y con clientes bien informados las empresas necesitan retener a sus clientes: un cliente no satisfecho puede expresar las razones de su abandono en Internet y las redes sociales dañando la imagen de la empresa y disminuyendo su probabilidad de captar nuevos clientes en el futuro y de retener los actuales. Comprender la perspectiva de los clientes y su grado de satisfacción es clave para la supervivencia de la empresa y la mejor

estrategia de *marketing* es la centrada en retener a los clientes y evitar su fuga (*customer churn*, en inglés) a la competencia (Kim y Yoon, 2004).

Existen dos tipos principales de estrategias ante el abandono de clientes: la conservadora y la proactiva. Una estrategia conservadora espera a que los clientes manifiesten su deseo de abandonar para ofrecerles incentivos para quedarse, como rebajas de tarifas o servicios especiales. La segunda, que es más efectiva, es identificar clientes descontentos y ofrecerles incentivos para quedarse, antes de que decidan abandonar. Para ello, es imprescindible establecer un sistema para reconocer a los clientes que comienzan a estar descontentos y estimar sus probabilidades de deserción. Es posible, entonces, diseñar una buena estrategia acoplando los incentivos a las probabilidades de abandono de manera que se maximice el beneficio esperado a largo plazo (Burez y Van den Poel, 2007).

El objetivo es disponer de una ecuación

$$p_i(t) = f(\mathbf{x}_i) \quad [2]$$

que nos permita prever la probabilidad de que un determinado cliente definido por un vector de características  $\mathbf{x}_i$  continúe utilizando los servicios de la compañía durante un periodo  $t$ . Las variables explicativas que pueden utilizarse son su historial de compras y relaciones con la empresa (quejas o reclamaciones, en su caso) y sus características personales. Para poder estimar esta ecuación necesitamos una base de datos de clientes que han permanecido fieles y, también, de exclientes que han abandonado el sistema. Estos datos son más fáciles de recoger en sectores donde el cliente y su actividad de compra de productos o servicios están bien identificados, como en las compañías telefónicas o de servicios de ocio por Internet, en empresas financieras etc., y la mayoría de los estudios realizados corresponden a estos sectores. Sin embargo, progresivamente otras empresas han tratado de recoger esta información buscando la fidelización de los clientes con tarjetas-cliente para descuentos, como las que han introducido muchos supermercados y superficies comerciales.

El procedimiento inicialmente utilizado para estimar una ecuación de lealtad del tipo [2] con pocas variables es el denominado modelo logístico, que hemos descrito brevemente en la sección 2 y cuya respuesta para una única variable explicativa se ha representado en la figura 3. Supongamos, por ejemplo, que la única variable explicativa,  $x$ , corresponde al consumo diferencial del cliente de un sistema de pago TV en un periodo determinado y que el consumo diferencial se calcula por la diferencia entre el tiempo de uso del cliente y el tiempo medio de uso de este sistema del cliente promedio. Si la relación entre la probabilidad de permanencia y el uso diferencial se explica por un modelo logístico la figura 3 indica que la probabilidad de permanencia es tanto mayor cuanto mayor sea su tiempo relativo de utilización. Con muchas variables la interpretación para cada una es similar y llamando  $p_i$  a la probabilidad de permanencia, y  $1 - p_i$  a la de abandono, la ecuación de un modelo logístico es del tipo:

$$g_i = \log \frac{p_i}{1 - p_i} = \beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki} \quad [3]$$

La ecuación [3] nos dice que en un modelo logístico el efecto de las variables sobre el logaritmo de la ratio entre las probabilidades de continuar y abandonar es lineal. Como comentamos en la sección 2, podemos estimar relaciones más complicadas de forma no lineal utilizando redes neuronales, que pueden considerarse la generalización no lineal de un modelo logístico (véase Caro y Peña [2021], para una explicación simple de estas redes para la predicción, y Peña y Tsay [2021] para un análisis más amplio aplicado a variables que evolucionan con el tiempo). Alternativamente, podemos utilizar un sistema de clasificación, como explicamos en la sección 2. Verbeke *et al.* (2011) han abordado este problema con máquinas de vector soporte que proporcionan el plano de separación entre las dos categorías de clientes y permiten también identificar las variables más importantes para la predicción.

En muchos casos la base de clientes de que disponemos no contiene siempre la información de si el cliente ha abandonado o no el sistema. Podemos entonces utilizar un sistema semisupervisado, donde tenemos la información completa para unos clientes, pero nos faltan las etiquetas para los otros. Esto supone realizar el estudio en dos pasos: primero clasificamos a todos los clientes sin etiqueta en uno de los dos grupos y luego, tomando como resultado la probabilidad asignada a cada uno, obtenemos la ecuación para el cliente de interés. En la sección siguiente describimos con más detalle el procedimiento de estimación de estas probabilidades en un sector muy competitivo, con muchas ofertas y productos similares: el sector de los grandes centros de distribución y venta de alimentos.

#### 4.1. Tipos de clientes y probabilidades de abandono

Los tipos de clientes pueden clasificarse por su frecuencia o intensidad de uso y por su actitud ante la empresa. Dick y Basu (1994), en un influyente artículo, propusieron cuatro tipos de clientes resultado de cruzar su intensidad de uso con la actitud ante la empresa. De esta forma podemos tener clientes leales (alta frecuencia y buena actitud), con lealtad latente (baja frecuencia, buena actitud), con lealtad espuria (alta frecuencia, baja actitud) y no leales (baja frecuencia y actitud). Heiens y Pleshko (1996) han estudiado la frecuencia de estos cuatro tipos de clientes en el mercado de hamburguesas de comida rápida encontrando una frecuencia similar en todas ellas, con la categoría más frecuente, no lealtad, del 32 % y las menos frecuentes, las lealtades espuria (20 %) y latente (18 %). Los leales representaban el 26 %.

Muchos estudios han relacionado la satisfacción del cliente con la lealtad y otros han puesto el énfasis en las relaciones emocionales y de identificación del cliente con la marca. Por ejemplo, Yu y Dean (2001) argumentan que la contribución de la satisfacción emocional es mejor predictor de la lealtad que la satisfacción cognitiva o racional. Knox y Denison (2000) han estudiado la lealtad al supermercado en el Reino Unido y los tipos de comportamiento del consumidor.

En el campo específico de estimación de probabilidad de abandono en un supermercado, Galeano y Peña (2019b) describen un ejemplo aplicado a una cadena de supermercados en España. Para detectar cuándo un cliente muestra signos prematuros de abandono, como

disminuir su frecuencia de compra o su intensidad de gasto, utilizan la existencia de una tarjeta de descuentos con la que se recoge información inmediata sobre sus compras. Con estos datos es posible generar para cada cliente una serie de actividad, que tiene el valor uno cuando el cliente realiza una compra o devolución y una serie de importes que recoge el coste de las compras, o devoluciones realizadas. Suponiendo que se elige un periodo fijo de análisis de un mes, en ese periodo para cada cliente  $i$  existe una probabilidad de actividad  $p_i$ . Esta probabilidad puede estimarse como el cociente entre los meses con actividad y los meses totales y clasificar a los clientes por su actividad. Típicamente, en los supermercados se detectan tres tipos de clientes:

- Clientes que tienen actividad todos los meses, o *clientes fieles*. La proporción de clientes fieles depende de cada supermercado pero varía entre el 10 % y el 30 % de los clientes totales. Además estos clientes son los que más gastan, del orden de un 100 % más que los frecuentes y del 150 % más que los ocasionales.
- Clientes con mucha actividad, o *clientes frecuentes*. Los clientes frecuentes son los que tienen una frecuencia de actividad menor del 100 %, pero superior a la mediana. Suelen representar entre el 20 % y el 50 % de los totales.
- Clientes con poca actividad, o *clientes ocasionales*. Los clientes ocasionales son los que tienen una frecuencia menor o igual a la mediana. Suelen representar entre el 40 % y el 60 % del total.

Existe evidencia (véase, por ejemplo, Galeano y Peña, 2019a) que en el sector de la alimentación el abandono se produce de forma gradual, de manera que se comienza probando la compra en otro centro que, si presenta ventajas competitivas, va sustituyendo poco a poco al actual. Por tanto, aunque por supuesto existen casos de abandono brusco por problemas concretos o cambios de domicilio, el abandono habitual de un cliente se manifiesta prematuramente en la disminución de la frecuencia de compra o de la intensidad de gasto en ella. Su comportamiento depende del tipo de cliente: un cliente fiel puede dejar de comprar todos los meses, pasando a la categoría de clientes frecuentes o puede disminuir el importe de sus compras, continuando como cliente fiel. Si deja de comprar en un mes dejará de pertenecer a la clase de fieles y será estudiado como cliente frecuente. Por tanto, para clientes fieles el único identificador de un posible abandono es una disminución en la magnitud de sus compras. Para los clientes frecuentes y ocasionales una disminución de actividad puede traducirse en un aumento del tiempo entre compras, un aumento de las rachas sin actividad, o en una disminución de los importes de compra.

Se puede construir un modelo que estime la probabilidad de abandono en función de su historia pasada de compras y de las variables que describen sus características personales. Su historial de compras puede representarse con variables como el nivel relativo actual de compras respecto al pasado, el número de cambios de nivel, positivos y negativos que han ocurrido en sus compras, su frecuencia de compra etc. Las variables personales pueden ser el sexo, la edad, el número de personas en el hogar, la antigüedad como cliente, el tipo de medios de pago utilizado, etcétera.

## 5. ANÁLISIS DE LOS CLIENTES EN RED

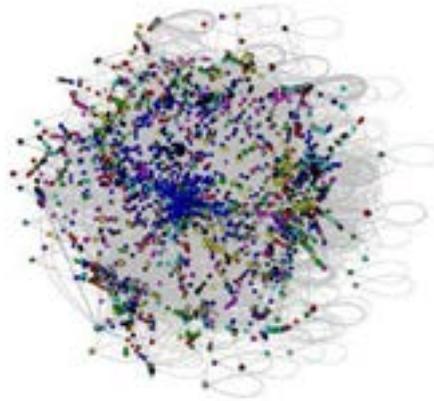
Una estructura potencialmente muy útil para el *marketing* digital es la consideración de los clientes en red. Una red, o grafo, puede representarse como un objeto matemático formado por vértices, o nodos, y aristas, o flechas. En su aplicación al *marketing* digital los nodos representan habitualmente clientes y las aristas transacciones o relaciones entre ellos. Por ejemplo, una red social, como Facebook o Instagram, está formada por nodos, que son las personas con cuenta en la red, y aristas, que representan las conexiones por relaciones de amistad entre las personas. Estas conexiones pueden ser todas iguales o podemos diferenciarlas, indicando en cada arista un número que representa la fuerza o intensidad de la relación. Por ejemplo, la intensidad puede definirse por el número de interacciones en la web entre las dos personas en un período de tiempo.

Una red de clientes se construye con la información disponible sobre las relaciones entre ellos. Por ejemplo, en un club deportivo o en un gimnasio se establecerán conexiones entre los usuarios que realizan actividades conjuntas. En una institución financiera podemos estar interesados en dos redes distintas: primero, de relaciones económicas entre los clientes; segundo, de relaciones personales. Cada cliente (que puede ser una persona o una empresa) será un nodo o vértice y las aristas representan relaciones o flujos entre dichos clientes, que suelen ser de distinta intensidad. Las relaciones pueden ser de parentesco o económicas. Por ejemplo, ser avalista de un crédito establece una relación económica entre los clientes, mientras que una transferencia recibida con un mensaje de felicitación indica, además, una relación de amistad o parentesco entre las dos personas. La representación gráfica de una red es útil con pocos clientes, véase por ejemplo la red representada en la figura 9 que comentaremos después, pero se vuelve muy compleja cuando este número es grande. Por ejemplo, la figura 6 representa una red de clientes en una institución bancaria con varios miles de personas y donde los colores indican distintos tipos de clientes. Es claro que el análisis de esta red requiere herramientas analíticas para su comprensión.

Para realizar un análisis descriptivo de una red se utilizan varias características, incluyendo medidas que midan la centralidad (importancia) de los clientes con el objetivo de cuantificar las relaciones de poder, protagonismo, confianza, etc., y la detección de comunidades específicas que puedan tener características interesantes a estudiar. Se denomina componente conexas de una red a una subred, es decir, una parte de la red, en la que todos sus vértices están conectados y a la que no se pueden añadir más vértices que cumplan dicha propiedad. Si un grafo tiene solo un componente conexo que incluye a todos los miembros, eso indica que siempre es posible llegar de un cliente a otro a través de un camino en la red. Si la red tiene varias componentes conexas puede dividirse en diversas redes separadas entre sí y donde los clientes no se comunican con los de otras redes. Habitualmente en redes de clientes la mayor componente conexas incluye una fracción muy significativa del mismo y se la denomina componente gigante. El resto de componentes se denominan componentes aisladas.

Un aspecto importante de las redes es obtener medidas de centralidad de los clientes. Por ejemplo, las figuras 7 y 8 representan distintos subgrupos en una red que no forman

Figura 6.

**Una red con muchos clientes**

*Fuente:* Elaboración propia.

componentes conexas porque están conectados unos con otros por clientes claves que hacen de puente entre estos subgrupos de clientes. En la figura 7 los grupos están identificados por colores distintos y hay varios clientes que juegan este papel de puente entre los subgrupos. Sin embargo, en la figura 8 se representan dos grandes grupos de clientes que solo están conectados por un nodo central que representa, en consecuencia, un cliente de gran valor estratégico.

Las medidas de importancia de los miembros de una red pueden calcularse de varias formas. Un primer enfoque es juzgar la importancia de un cliente por el número de relaciones o vértices que llegan a él y, si estas tienen distinta intensidad, ponderadas por su importancia relativa. Este método tiene la ventaja de que permite calcular la importancia de cada cliente aisladamente, y el inconveniente es que solo utiliza la información de las relaciones directamente conectadas al cliente. Una segunda opción más compleja es definir la centralidad de un cliente por la importancia media de los clientes a los que está conectado. Este método utiliza toda la información de la red para definir la importancia de cada cliente, pero requiere obtener todas las importancias de los clientes simultáneamente.

Este segundo enfoque fue el utilizado por Google para definir la importancia de las páginas web en su buscador (Page *et al.*, 1999). Para explicar su construcción, consideremos una red con  $N$  vértices de clientes y definamos la importancia del cliente  $i$ , que llamaremos  $I(i)$ , como un promedio de la importancia de los clientes a los que está conectado, es decir:

$$I(i) = c \sum_{j \rightarrow i} I(j) / N_i \quad [4]$$

donde el sumatorio afecta a todos los vértices que están conectados con el vértice  $i$  y  $N_i$  es el número total de vértices que están conectados al vértice  $i$ . Tendremos una ecuación con

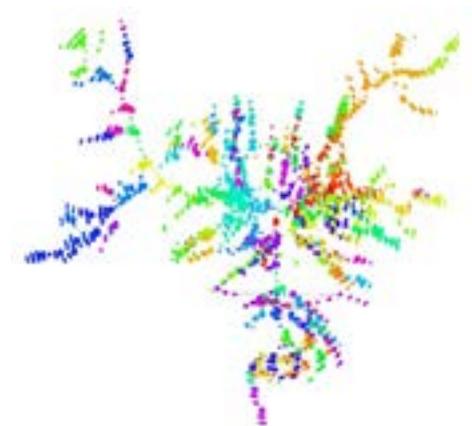
esta forma para cada uno de los  $N$  clientes en la red. Para resolver este sistema de ecuaciones llamemos  $I$  al vector con  $N$  componentes que representan la importancia de cada uno de los  $N$  vértices de la red. Definamos ahora una matriz  $A$  de conexiones entre vértices, que tiene coeficientes  $a_{ij}$  que tomarán el valor 0 si los nudos  $i$  y  $j$  no están conectados y el valor  $1/N_i$  cuando lo están. Es decir, la fila de la matriz  $i$  que corresponde a este vértice tiene  $N - N_i$  elementos iguales a cero, para los vértices no conectados al  $i$ , y  $N_i$  vértices para los conectados al  $i$ , con valores  $1/N_i$ . La suma de los elementos de las filas de la matriz es, por tanto, siempre igual a uno, es decir, llamando  $\mathbf{1}$  al vector de unos se verifica que el producto de la matriz de conexiones,  $A$ , por el vector de unos,  $\mathbf{1}$ , es otro vector de unos, es decir:

$$A\mathbf{1} = \mathbf{1} \quad [5]$$

Esta condición implica que el vector de unos no se modifica al multiplicarlo por la matriz  $A$ . Se denominan vectores característicos o propios de una matriz aquellos que solo se modifican por una constante al multiplicarlos por la matriz.

Figura 7.

**Una red con varios grupos de clientes (los colores indican diferentes tipos de clientes)**



Fuente: Elaboración propia.

El conjunto de las  $N$  ecuaciones [4] puede escribirse utilizando la notación de vectores y matrices como:

$$I = cAI \quad [6]$$

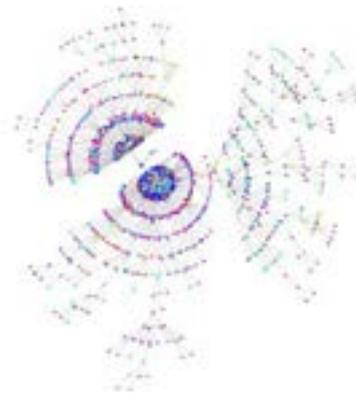
y el vector de importancias que buscamos,  $I$ , debe ser un vector propio de la matriz de conexiones  $A$ , multiplicado por una constante, que se denomina valor propio,  $c$ . Esta ecuación tiene la solución trivial  $I = \mathbf{1}$  y  $c = 1$  donde todos los vértices tienen la misma importancia y corresponde a la condición que hemos establecido de que las filas de la matriz de conexiones

sumen la unidad. La solución que buscamos es aquella que hace máxima  $c < 1$  y el vector propio asociado será el que define la importancia de cada vértice. Esta solución puede obtenerse con un método iterativo general para obtener vectores y valores propios de una matriz. Este criterio se conoce a veces como la centralidad por valor propio o centralidad Pagerank, y fue desarrollado por Page *et al.* (1999) para medir la importancia de las páginas web en el buscador inicial desarrollado por Google. Las páginas de la web forman una red donde cada una está conectada a otras por links y la importancia de una página depende de la importancia de sus conexiones a otras. Este método ha sido utilizado extensamente a partir del éxito alcanzado por el buscador de Google.

La red de clientes de una empresa es una fuente de información muy importante para el *marketing* digital en varios aspectos. En primer lugar, permite entender y medir la importancia de cada cliente analizando su posición en dicha red. Tradicionalmente, se ha medido la importancia de un cliente por el valor de sus compras o de sus activos. Sin embargo, una mejor medida es el coste que supone su pérdida. Un cliente aislado deja de serlo y las consecuencias negativas son su retirada. La deserción de un cliente muy conectado puede producir consecuencias negativas por su influencia negativa sobre otros clientes. El coste de perder un cliente depende de su importancia estratégica dentro de la red de clientes, que puede hacer que esta pérdida se multiplique.

Figura 8.

### **Dos grupos de clientes conectados por una única persona**



Fuente: Elaboración propia.

Por ejemplo, consideremos el efecto de una persona muy influyente en las redes sociales que envía informaciones poco favorables sobre la empresa. El coste de su marcha puede ser mucho mayor que la pérdida de los ingresos que aportaba, si su comportamiento arrastra a otros usuarios a seguir su ejemplo. Conociendo las relaciones entre los clientes en forma de red, podemos medir la importancia de cada una de forma mucho más precisa. Por ejemplo,

en la figura 8, dos grandes grupos de clientes están conectados por una persona. Si esta desaparición de las consecuencias de su marcha pueden ser muy importantes.

En segundo lugar, la red de clientes es una pieza fundamental para captar la presencia de grupos de clientes cohesionados (con fuertes relaciones entre ellos). Conocer la presencia de estos grupos es muy relevante para orientar determinadas políticas comerciales. Por ejemplo, podemos analizar las características de estos grupos y crear productos específicos apropiados para ellos, como en los sistemas de recomendación. Sin embargo, la ventaja de la red es poder localizar a los clientes centrales dentro de la subred creada por cada grupo para mejorar la distribución de la información sobre dichos productos. Determinar la presencia de grupos de clientes fuertemente enlazados se conoce con el nombre de detección de comunidades. El problema consiste en obtener una partición de la red en varias subredes, de tal manera que los miembros de cada una de las subredes tengan una fuerte relación entre ellos y escasa con los miembros de otras. Una medida de la calidad de una partición es la modularidad, que se define como la suma para cada una de las subredes de las diferencias entre el número de relaciones dentro de la subred con respecto al esperado si no existiese tal estructura de comunidades. Cuanto mayor sea el valor de la modularidad de una red, más significativa será la presencia de los grupos.

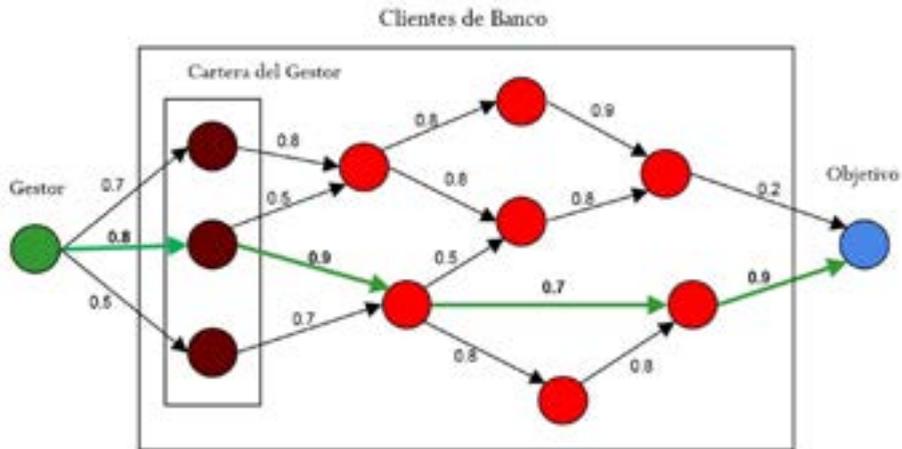
En tercer lugar, la red de clientes puede ser utilizada para explicar y predecir determinados comportamientos de los clientes y prever las consecuencias de sucesos económicos. Por ejemplo, en una red de clientes de una institución bancaria podemos estudiar el efecto de que una empresa que paga sus nóminas a sus trabajadores, que son miembros de la red, entre en suspensión de pagos. Esto desencadena un efecto que puede seguirse a lo largo de la red y que afecta a sus proveedores, pero también a muchos otros, como por ejemplo a los propietarios de viviendas alquiladas por los trabajadores de esa empresa.

En cuarto lugar, las variables de la red pueden ser útiles para explicar otras variables de interés. Por ejemplo, Galeano y Peña (2019a) y (2019b) detectaron que la red de clientes de un banco contiene variables importantes para la predicción de la morosidad. Prever que un cliente va a entrar en mora es *a priori* un problema complicado. Una opción es utilizar su información personal, como su edad, su estado civil, su salario y patrimonio etc. Sin embargo, se detectó que la variable más importante para prever la morosidad es el número de relaciones a un paso que tiene el cliente en la red con otros clientes que han entrado en mora. *A posteriori* este hecho parece muy razonable, ya que si una persona tiene sus relaciones con personas sin problemas económicos es más fácil suponer que tampoco los tendrá. Sin embargo, estas variables solo pueden medirse cuando se construye la red.

En quinto lugar la red puede ser importante para desarrollar nuevas políticas comerciales. Por ejemplo, las relaciones entre los clientes pueden ayudar a los comerciales a captar o influir en otros. Como ejemplo, la figura 9 muestra un ejemplo de una red diseñada para captar clientes a través de gestores que utilizan las relaciones de los clientes para conectar con otros. Los valores que aparecen en las aristas son las probabilidades de éxito. Véase Quijano-Sánchez y Liberatore (2017) para una descripción de este tipo de redes en el sector financiero.

Figura 9.

**Ejemplo de una red para captar clientes (los números en las aristas representan la probabilidad de que la interacción del nudo origen al destino representada por la arista que los une tenga éxito)**



Fuente: Elaboración propia.

Por último, se está iniciando el estudio de las redes dinámicas, donde las relaciones entre los nodos se representan como series temporales, es decir, con una secuencia de valores a lo largo del tiempo. Por ejemplo, en una red social podemos representar la relación mensual entre dos personas por la secuencia de sus intercambios diarios, que forman una serie temporal. El análisis de la red implica ahora entender cómo unas relaciones influyen sobre otras a lo largo del tiempo, lo que permitirá generar predicciones futuras.

## 6. CONCLUSIONES

Las ventas a través de Internet tendrán una importancia creciente en el futuro, impulsando las herramientas de *marketing* digital basadas en la información recogida sobre los clientes. Además, incorporarán cada vez más datos obtenidos de audios y vídeos, y, en algunos sectores específicos, provenientes de otros sentidos digitalizados, como el olor, en perfumería, el gusto, en alimentación, y el tacto, en telas, muebles o vestidos.

En este trabajo hemos comentado tres herramientas de *marketing* digital basadas en los datos masivos, pero su disponibilidad y creciente desarrollo tendrá un efecto muy amplio. Por ejemplo, en la detección del fraude, que es un problema central para hacer las compras seguras. Disponer de la historia completa de los pagos con tarjeta de una persona puede hacer

muy fácil la identificación de una operación fraudulenta y avisar al usuario de forma inmediata. Además, permitirá clasificar los tipos de fraude más habituales y prevenirlos, con lo que la seguridad de las actividades en Internet puede experimentar mejoras muy importantes (véase Galeano y Peña, 2019b).

La transformación digital que estamos experimentando abarca muchas dimensiones (Kanan, 2017): afecta al comportamiento de los consumidores, a la información disponible para la compra, a las plataformas que pueden utilizarse y a la interacción entre estos aspectos y el contexto del cliente. Por ejemplo, el comportamiento del consumidor digital respecto a la información ha cambiado. En primer lugar, su opinión se forma no solo con la que publicita la empresa, sino, también, con la experiencia de otros usuarios que comparten opiniones en Internet. En segundo lugar, el mismo cliente se convierte en generador de valoraciones de un producto o servicio al comentar su experiencia en las redes sociales y webs de intercambio de opiniones de consumidores. También su experiencia de compra se ha transformado. Ahora es posible comprar en el marco tradicional de empresa/cliente pero también en el de relaciones cliente/cliente, como por ejemplo en Ebay y otras plataformas. Además, grandes intermediarios, como Amazon, conectan al cliente con muchas empresas, formando relaciones cliente/empresas. Finalmente, aunque los productos se ofrezcan a un mercado mundial su venta digital está afectada por fenómenos de piratería o de regulaciones nacionales, que varían mucho con el contexto geográfico y cultural.

La comprensión de todos estos fenómenos requiere el análisis de los datos que los describen, que se van acumulando a gran velocidad a medida que se generalizan las transacciones comerciales por Internet, generando grandes bases de datos masivos sobre todos estos aspectos. Su conversión en información mediante tratamientos analíticos va a ofrecer enormes posibilidades para perfeccionar las estrategias comerciales de *marketing* digital en el futuro y hacer nacer otras nuevas.

## Referencias

- BENITO, M., GARCÍA-PORTUGUÉS, E., MARRON, J. S. y PEÑA, D. (2017). Distance-weighted discrimination of face images for gender classification. *Stat*, 6(1), pp. 231-240.
- BENITO, M. y PEÑA, D. (2005). A fast approach for dimensionality reduction with image data. *Pattern Recognition*, 38(12), pp. 2400-2408.
- BRUCE, V. (1988). *Recognising Faces, Essays in cognitive psychology*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- BRUCE, V., BURTON, A. M. y CRAW, I. (1992). Modelling face recognition. *Philosophical Transactions of the Royal Society, Series B*, 335(1273), pp. 121-127.
- BUREZ, J. y VAN DEN POEL, D. (2007). CRM at a pay-TV company: Using analytical models to reduce customer attrition by targeted marketing for subscription services. *Expert Systems with Applications*, 32(2), pp. 277-288.
- CARO, Á. y PEÑA, D. (2021). Predicción de series temporales económicas con datos masivos: perspectiva, avances y comparaciones. En: D. PEÑA, P. PONCELA y E. RUIZ (eds.), *Nuevos métodos de Predicción Económica con datos masivos*. Madrid: Funcas. <https://www.funcas.es/articulos/prediccion-de-series-temporales-economicas-con-datos-masivos-perspectiva-avances-y-comparaciones/>

- CHAFFEY, D. y ELLIS-CHADWICK, F. (2019). *Digital marketing*. UK: Pearson.
- DICK, A. S. y BASU, K. (1994). Customer loyalty: toward an integrated conceptual framework. *Journal of the Academy of Marketing Science*, 22(2), pp. 99-113.
- EREVELLES, S., FUKAWA, N. y SWAYNE, L. (2016). Big Data consumer analytics and the transformation of marketing. *Journal of Business Research*, 69(2), pp. 897-904.
- GALEANO, P. y PEÑA, D. (2019a). Data science, big data and statistics. *Test*, 28(2), pp. 289-329.
- GALEANO, P. y PEÑA, D. (2019b). Las nuevas oportunidades del Big Data para las instituciones financieras. *Papeles de Economía Española*, 78(162), pp. 78-176. Madrid: Funcas. <https://www.funcas.es/articulos/las-nuevas-oportunidades-de-big-data-para-las-instituciones-financieras-la-gestion-de-la-informacion-en-banca-de-las-finanzas-del-comportamiento-a-la-inteligencia-artificial-2019-n-162/>
- GALTON, F. (1910). Numeralized profiles for classification and recognition. *Nature*, 83, pp. 127-130.
- GONG, Y. (1995). Speech recognition in noisy environments: A survey. *Speech communication*, 16(3), pp. 261-291.
- HEIENS, R. A. y PLESHKO, L. P. (1996). Categories of customer loyalty: An application of the customer loyalty classification framework in the fast food hamburger market. *Journal of Food Products Marketing*, 3(1), pp. 1-12.
- HULTÉN, B. (2017). Branding by the five senses: A sensory branding framework. *Journal of Brand Strategy*, 6(3), pp. 281-292.
- HUSSAIN, S. (2019). Sensory Marketing Strategies and Consumer Behavior: Sensible Selling Using All Five Senses. *IUP Journal of Business Strategy*, 16(3).
- KANNAN, P. K. (2017). Digital marketing: A framework, review and research agenda. *International Journal of Research in Marketing*, 34(1), pp. 22-45.
- KATZENBEISSER, S. y PETKOVIC, M. (2008). Privacy-preserving recommendation systems for consumer healthcare services. En: *2008 Third International Conference on Availability, Reliability and Security*, pp. 889-895.
- KAWULOK, M., CELEBI, ME and SMOLKA, B (2016). *Advances in Face Detection and Facial Image Analysis*, Springer Publishing Company, Inc.
- KHAN, K., ATTIQUE, M., KHAN, R. U., SYED, I. y CHUNG, T. S. (2020). A multi-task framework for facial attributes classification through end-to-end face parsing and deep convolutional neural networks. *Sensors*, 20(2), p. 328.
- KIM, H. S. y YOON, C. H. (2004). Determinants of subscriber churn and customer loyalty in the Korean mobile telephony market. *Telecommunications Policy*, 28(9-10), pp. 751-765.
- KNOX, S. D. y DENISON, T. J. (2000). Store loyalty: its impact on retail revenue. An empirical study of purchasing behaviour in the UK. *Journal of Retailing and Consumer Services*, 7(1), pp. 33-45.
- KUMAR, P. y THAKUR, R. S. (2018). Recommendation system techniques and related issues: A survey. *International Journal of Information Technology*, 10(4), pp. 495-501.
- LEVI, G. y HASSNER, T. (2015). Age and gender classification using convolutional neural networks. En: *Proceedings of the IEEE conference on computer vision and pattern recognition workshops* (pp. 34-42).
- LINDEN, G., SMITH, B. y YORK, J. (2003). Amazon.com Recommendations: Item-to-Item Collaborative Filtering. *IEEE Internet Computing*, vol. 7(1), pp. 76-80.
- MARX, V. (2013). The big challenges of big data. *Nature*, 498(7453), pp. 255-260.
- NASSIF, A. B., SHAHIN, I., ATTILI, I., AZZEH, M. y SHAALAN, K. (2019). Speech recognition using deep neural networks: A systematic review. *IEEE access*, 7, pp. 19143-19165.
- PADMANABHAN, J. y JOHNSON PREMKUMAR, M. J. (2015). Machine learning in automatic speech recognition: A survey. *IETE Technical Review*, 32(4), pp. 240-251.

- PAGE, L., BRIN, S., MOTWANI, R. y WINOGRAD, T. (1999). The PageRank citation ranking: Bringing order to the web. *Stanford InfoLab*.
- PEÑA D. (2015). Big data, Ciencia y Estadística. *Revista de Ciencia y Humanidades*, 14, pp. 97-106. Fundación Ramón Areces.
- QUIJANO-SANCHEZ, L. y LIBERATORE, F. (2017). The BIG CHASE: A decision support system for client acquisition applied to financial networks. *Decision Support Systems*, 98, pp. 49-58.
- RYAN, D. (2016). *Understanding digital marketing: marketing strategies for engaging the digital generation*. Kogan Page Publishers.
- SAURA, J. R. (2021). Using data sciences in digital marketing: Framework, methods, and performance metrics. *Journal of Innovation and Knowledge*, 6(2), pp. 92-102.
- SAYYAD, S. et al. (2018). Digital Marketing Framework Strategies Through Big Data. En: *International conference on Computer Networks, Big data and IoT*, pp. 1065-1073. Cham: Springer.
- SHAN, C. (2012). Learning local binary patterns for gender classification on real-world face images. *Pattern Recognition Letters*, 33(4), pp. 431-437.
- SMITH, B. y LINDEN, G. (2017). Two decades of recommender systems at Amazon. *IEEE Internet Computing*, 21(3), pp. 12-18.
- TRAN, S., DU, M., CHANDA, S., MANMATHA, R. y TAYLOR, C. (2019). Searching for Apparel Products from Images in the Wild. arXiv preprint arXiv:1907.02244.
- VELASCO, C., WAN, X., KNOEFERLE, K., ZHOU, X., SALGADO-MONTEJO, A. y SPENCE, C. (2015). Searching for flavor labels in food products: the influence of color-flavor congruence and association strength. *Frontiers in Psychology*, 6, p. 301.
- VERBEKE, W., MARTENS, D., MUES, C. y BAESENS, B. (2011). Building comprehensible customer churn prediction models with advanced rule induction techniques. *Expert Systems with Applications*, 38(3), pp. 2354-2364.
- WANG, Y. F., CHIANG, D. A., HSU, M. H., LIN, C. J. y LIN, I. L. (2009). A recommender system to avoid customer churn: A case study. *Expert Systems with Applications*, 36(4), pp. 8071-8075.
- WEDEL, M. y KANNAN, P. K. (2016). Marketing analytics for data-rich environments. *Journal of Marketing*, 80(6), pp. 97-121.
- YU, Y. T. y DEAN, A. (2001). The contribution of emotional satisfaction to consumer loyalty. *International journal of service industry management*.
- ZHAO, W., CHELLAPPA, R., PHILLIPS, P. J. y ROSENFELD, A. (2003). Face recognition: a literature survey. *ACM Computing Surveys*, 35(4), pp. 399-458.