

Social Media Analytics

Àlex Bartrolí Muñoz

PID_00209852



Los textos e imágenes publicados en esta obra están sujetos –excepto que se indique lo contrario– a una licencia de Reconocimiento-NoComercial-SinObraDerivada (BY-NC-ND) v.3.0 España de Creative Commons. Podéis copiarlos, distribuirlos y transmitirlos públicamente siempre que citéis el autor y la fuente (FUOC. Fundación para la Universitat Oberta de Catalunya), no hagáis de ellos un uso comercial y ni obra derivada. La licencia completa se puede consultar en <http://creativecommons.org/licenses/by-nc-nd/3.0/es/legalcode.es>

Índice

Introducción	5
1. BI en redes sociales	7
1.1. Analítica de negocio en la web antes de las redes sociales	8
1.2. Tipos de redes sociales	9
1.3. Ventajas de las redes sociales para el BI	10
1.3.1. Segmentación de mercado	12
1.3.2. Mejora de la atención al cliente	12
1.3.3. Predicciones de mercado	13
1.3.4. Estudio de la competencia	14
1.3.5. Acceso a nichos de mercado	14
1.3.6. Estudios de mercado	14
1.3.7. Comunicación inter e intraempresas	15
1.4. Uso de datos móviles en sistemas de BI	15
2. Métricas e indicadores	17
2.1. Métricas en redes sociales	17
2.2. Indicadores	18
3. Herramientas de medición	22
3.1. Herramientas de medición de datos de redes sociales	22
3.2. Ejemplo: Facebook Insights	23
3.3. Herramientas de medición de dispositivos móviles	25
3.4. Ejemplo: Yellowfin Business Intelligence	26
4. Ejemplo de BI en redes sociales	30
5. Almacenamiento y recuperación de datos de redes sociales ...	32
5.1. Recuperación de información de redes sociales	32
5.1.1. <i>Plugins</i> de las redes sociales	33
5.1.2. API de las redes sociales	33
5.2. Bases de datos para datos procedentes de redes sociales	34
6. Técnicas y algoritmos para la clasificación y personalización en social BI	37
6.1. Agrupaciones de individuos	37
6.2. Clasificación de individuos	39
6.2.1. Ejemplo de clasificador de la intención de usuario	40
6.3. Similitud entre elementos	42
6.3.1. Distancia euclídea	43
6.3.2. Coeficiente de Sorensen-Dice y distancia de Jaccard	43

6.3.3. Distancia de Hamming	45
6.3.4. Distancia de Levenshtein	45
6.4. Teoría de grafos	45
7. Sistemas de recomendación.....	47
7.1. Ejemplo de sistemas de recomendación de películas	49
Resumen.....	53
Bibliografía.....	55

Introducción

En el transcurso de los últimos años cada vez son más las empresas que utilizan soluciones de Business Intelligence (BI) para la toma de decisiones que mejoren sus resultados empresariales y los productos y servicios ofrecidos a sus clientes. Debido a ello, las empresas se están focalizando en varios aspectos clave que se consideran relevantes en una solución BI.

Con el objetivo de entender mejor a sus clientes, las empresas están incorporando datos de redes sociales y de dispositivos móviles en sus sistemas de BI, permitiendo analizar aspectos de negocio hasta ahora inalcanzables. Dado que el BI debe servir para la toma de decisiones, la información se tiene que presentar de forma resumida, a ser posible gráficamente, y mostrando solo aquellos elementos relevantes que necesitan interpretación. Para ello, aunque se utilicen datos de las redes sociales en el proceso de BI, se sigue trabajando con herramientas de BI que sintetizan la información y muestren cuadros de mando efectivos para la toma de decisiones.

Las analíticas que deberán mostrar estas herramientas deberán ser proactivas y orientadas a acciones futuras, y tendrán que explicar de forma esquemática y sintetizada el histórico de hechos o acciones que han influido en la situación actual. Asimismo, las decisiones se deberían poder tomar con información en tiempo real, sin necesidad de tener que esperar. Esta necesidad de tiempo real es aún más patente cuando se analizan datos de las redes sociales. Las redes sociales son muy dinámicas y masivas y, en consecuencia, reaccionar un día tarde puede representar grandes pérdidas a la empresa, tanto en clientes como en facturación o incluso en su reputación.

Con la inclusión de datos móviles y de redes sociales aparecen nuevas problemáticas. Una de ellas es que, para poder procesar todos los datos necesarios, las arquitecturas tradicionales pueden resultar insuficientes. Actualmente, las arquitecturas distribuidas y la computación en la nube (*cloud computing*) están cada día más presentes en las empresas y el almacenamiento de datos en bases de datos NoSQL es cada vez más común.

La primera parte de este módulo empieza describiendo las redes sociales y las ventajas que puede tener utilizar su información en los procesos analíticos de una empresa. En el segundo apartado se comentan las principales métricas utilizadas en las redes sociales y ejemplos de indicadores clave de desempeño (más conocidos como KPI) de alguna de las redes sociales más conocidas. A continuación, ya en el apartado 3, se muestran unos ejemplos de herramientas de medición de redes sociales y de Mobile BI. En el apartado 4 se explica una estrategia de social media llevada a cabo en Facebook y Twitter por una empresa de venta de helados para aumentar el conocimiento de su marca e

incrementar ingresos. Seguidamente, en el apartado 5, se muestran diferentes herramientas para mostrar o recuperar información de redes sociales y se comentan los sistemas de almacenamiento más utilizados en social BI. El apartado 6 está destinado a las herramientas estadísticas y matemáticas más utilizadas para el análisis de datos de redes sociales. Para finalizar, el apartado 7 describe cómo utilizar sistemas de recomendación que utilicen información extraída a partir de redes sociales.

1. BI en redes sociales

Durante los últimos años hemos asistido al nacimiento y crecimiento explosivo de las redes sociales. Estas, gracias a su facilidad de uso, han supuesto una revolución en la comunicación y han cambiado de forma significativa los patrones que teníamos de ella. Actualmente las redes sociales forman parte de nuestro día a día: compartimos nuestra vida social y nos entretenemos en Facebook, buscamos trabajo en LinkedIn y nos informamos con Twitter.

Esta nueva forma de comunicación ha traído otros cambios importantes:

- El volumen de información generada suele ser muy superior a la que se gestionaba anteriormente. Eso ha hecho que para poder tratarla eficientemente, en algunos casos sea necesario utilizar nuevos sistemas de almacenamiento, dejando de lado las bases de datos relacionales.
- El origen y formato de esta información es muy diverso. Hasta hace poco la información utilizada en medios *online* era mayoritariamente en formato texto, imagen y vídeo, y estaba relativamente estructurada. Con las redes sociales se debe añadir otro tipo de información, como los gustos u opiniones de las personas, que se pueden medir con los “Me gusta”; el impacto de una persona, medible a partir su número de seguidores; la importancia de una información, medible a partir del número de veces que se ha compartido o se ha hecho retuit y el nacimiento de un tema de interés o debate, medible con los *trending topics*.
- El intercambio de información es inmediato y global. Aunque ya había otros medios de comunicación donde la información se podía difundir en directo, como la televisión o la radio, las redes sociales y las tecnologías móviles han ampliado muchísimo su efecto, permitiendo que todo el mundo pueda difundir información y de cualquier tipo. Por ejemplo, en la temporada 2013-14, un jugador del F.C. Barcelona criticó, en su cuenta Twitter, la actuación arbitral de un partido del Madrid por no pitar un penalti claro en contra del conjunto blanco en los últimos minutos del partido. Este mensaje condicionó la rueda de prensa postpartido del entrenador madridista y la segunda pregunta que se realizó no fue sobre el partido, sino sobre el tuit emitido por el futbolista unos minutos antes.

Lectura recomendada

“Piqué tras el penalti al Real Madrid: ‘Viendo una película de humor en Canal + Liga’”. *elEconomista.es* (26/09/2013).

Los cambios introducidos por las redes sociales no se limitan solo a la comunicación entre personas, sino que influyen en muchos otros sectores de la sociedad y la economía. Algunos ejemplos de campos o actividades en donde pueden ser útiles son la segmentación de mercado, la creación de nuevos pro-

ductos, la medición de la popularidad, la atención al cliente, la predicción de tendencias en valores bursátiles, la personalización del contenido para usuarios, etc.

1.1. Analítica de negocio en la web antes de las redes sociales

Antes de que existieran las redes sociales, el número de usuarios en Internet era muy elevado, y los portales web con un elevado número de visitantes podían personalizar su contenido para mostrar información distinta a cada usuario, o bien modificar la apariencia de sus páginas en función de las páginas que visitaba cada usuario mediante técnicas de analítica web. A continuación se comentan brevemente los distintos mecanismos que se utilizaban para personalizar la información (y por lo tanto, servicios y productos) proporcionada al usuario.

Desde hace años se utilizan técnicas como la personalización por IP, el uso de *cookies* o de cuentas de usuario para personalizar el contenido que se muestra a cada usuario. Así, en webs como Amazon, en función de las compras realizadas por un usuario y de los productos vistos, el sistema es capaz de aprender de los gustos de los usuarios y, en consecuencia, personalizar los productos mostrados para cada usuario. Para ello no se basa solo en las opiniones del usuario, sino que utiliza la opinión de aquellos usuarios “vecinos”, entendiendo usuarios vecinos como aquellos usuarios con gustos similares. Por ejemplo, si un usuario U_X ha comprado los libros A, B, C, D y E y un usuario U_Y ha comprado los libros C, D, E, F y G, entonces la web considerará que los dos usuarios tienen gustos similares (tienen 3 libros en común) y por lo tanto son vecinos. En consecuencia, mostrará los libros F y G al usuario U_X y los libros A y B al usuario U_Y . Este tipo de personalización ha sido el más utilizado hasta el momento, debido a los buenos resultados obtenidos y a la facilidad de su puesta en marcha.

También se ha trabajado mucho con la personalización de contenido en webs de noticias y portales de contenido, personalizando las *landing pages* principales a partir de aquellas noticias que interesaban más a sus lectores. Así, noticias o tipos de noticias que inicialmente no ocupaban una posición en la parte más visible de las *landings pages* pasaron a posiciones más prominentes debido a la importancia que le daban los usuarios. Para realizarlo se utilizaban técnicas de análisis de visitantes similares a las que hoy en día ofrece Google Analytics. Después, la información mostrada se personalizaba en función de la región desde la que se conectaba el usuario.

En otro tipo de portales web, como los portales de contactos, la personalización se realiza mediante datos más explícitos. En estos casos los usuarios tienen que rellenar unos largos formularios para describir sus gustos y su personalidad. A partir de sus datos, el sistema identifica a las personas más afines.

Landing page

Una *landing page* es una página a la que acostumbra a llegar el usuario cuando visita la web por primera vez o cuando accede a ella de forma recurrente. Estas páginas, normalmente, muestran un listado de otras páginas (o productos en el caso de una tienda *online*) a las que el usuario se dirige desde estas páginas y suelen estar organizadas por temáticas (noticias de cultura, economía, deporte, portátiles, novedades, etc.) o bien ser la página principal de la web.

Los datos que podían extraerse de la web antes de la aparición de las redes sociales no eran tan relevantes como los que se pueden obtener hoy en las mismas. En las redes sociales los usuarios “hablan”, de forma crítica, de productos, de servicios, de marcas, de empresas, etc. Procesar y analizar esos datos para inferir información sobre la percepción de una empresa, o de sus servicios, por los potenciales clientes es un proceso que puede sustituir la información que anteriormente solo podía obtenerse mediante complejos y caros estudios de mercado.

1.2. Tipos de redes sociales

Cuando hablamos de redes sociales, la gente puede pensar en Facebook, Twitter, LinkedIn, Pinterest, Youtube y alguna más, pero en realidad existen muchas redes sociales y de muchos tipos, como podemos ver en la figura 1.

Figura 1. Tipos de redes sociales



Fuente: www.FredCavazza.net.

Las redes sociales, según su ámbito de aplicación, se pueden clasificar en:

1) **Redes sociales horizontales:** Redes sociales de ámbito general. No buscan un tipo de usuario concreto, sino que permiten la libre participación de quien así lo desee, proporcionando una herramienta de interacción a nivel general. Ejemplos de redes sociales horizontales son Facebook, Twitter, Google+, MySpace, Hi5, Sonico y Pinterest.

2) **Redes sociales verticales:** Son redes focalizadas a un ámbito particular. Esta clase de red social está dirigida a un público determinado como sería el profesional o de ocio. Entre las redes verticales destacan las redes sociales relacionadas con:

- Búsqueda de trabajo: LinkedIn, Viadeo y Xing.
- Geolocalización: son muy utilizadas en dispositivos móviles y acostumbran a ofrecer recomendaciones sobre actividades, experiencias o lugares de interés cercanos: centros culturales, restaurantes, lugares de ocio, etc. Destacan Foursquare, Bliquo y Facebook Places.
- Fotografía: Flickr y Fotolog.
- Vídeo: Vimeo y Youtube.
- Viajes: Tripadvisor, TravBuddy, Minube.
- Música: Last.fm, Gatunes, Mog.
- Libros: Librarything, Shelfari, Quelibroleo.
- Noticias: Reddit, Meneame.
- Juegos: Wipley, Popuz, Gamerate.
- Cine y TV: Flixster, GetGlue.
- Motos: Moterus, Eburra.
- Animales: Dogster, TV Mascotas.
- Turismo: Hosteltur, Turismo 2.0.
- Gastronomía: Yelp.

A partir de estas redes sociales se genera una cantidad de datos importantísima para conocer y relacionar mejor la sociedad en la que vivimos. Este gran volumen de datos, ha propiciado la utilización de técnicas de Big Data para su almacenamiento y análisis, pero también ha aportado una fuente de información de valor incalculable para las empresas que, gracias a soluciones BI, han visto cómo se les ha abierto una puerta de posibilidades impensable unos años atrás. A continuación vamos a ver algunas de las ventajas que el análisis de la información de redes sociales puede aportar a las empresas.

1.3. Ventajas de las redes sociales para el BI

La información obtenida de los usuarios de redes sociales es muy heterogénea y permite inferir gran cantidad y diversidad de información.

Al disponer de tanta información relevante, conseguida casi de forma instantánea y a un coste muy reducido, los procesos analíticos pueden ser mucho más exhaustivos y la toma de decisiones puede estar respaldada por indicadores más precisos.

Además, las redes sociales ofrecen a las empresas nuevas vías de trabajo y facilitan muchas otras. Permiten recuperar de forma centralizada información de clientes muy dispersos geográficamente, permitiendo así la creación de cuadros de mando e informes integrales, que pueden ser consultados al mismo tiempo por varios encargados para la toma de decisiones simultáneas en diferentes lugares. Podemos decir que, gracias a las redes sociales, muchas herramientas de toma de decisiones han pasado a ser herramientas 2.0.

En esta sección nos centraremos en las ventajas de las redes sociales en los procesos analíticos. La primera, y más evidente ventaja de las redes sociales para nuestra empresa, es el alto grado de conocimiento que podemos tener de nuestros clientes. De entre toda la información, podemos destacar que:

- Conocemos mejor sus gustos sobre productos a través de los “Me gusta” en su página de Facebook o similares.
- Obtenemos una estimación sobre sus temas de interés a partir de información sobre:
 - Las compras (por ejemplo de libros) que hacen por Internet, o de los vídeos o series que miran.
 - Las personas a las que siguen en Facebook o Twitter.
- Conocemos su grado de satisfacción con una empresa o producto a través de sus comentarios, en *blogs*, foros, Facebook o Twitter.

Además de un mejor conocimiento de nuestros clientes, existen también otras ventajas que se derivan del análisis de la información de redes sociales. Con ellas, de forma más económica y eficaz, podemos:

- Segmentar el mercado de manera más eficiente.
- Mejorar la atención al cliente.
- Realizar predicciones de mercado.
- Estudiar a la competencia.
- Acceder a nichos de mercado.
- Realizar estudios de mercado.
- Mejorar la comunicación inter e intraempresas.

En los siguientes apartados ampliamos las explicaciones sobre las ventajas que acabamos de enumerar.

1.3.1. Segmentación de mercado

La segmentación de mercado es el arte de dividir un mercado en distintos grupos de consumidores que puedan requerir productos o servicios diferentes. Para llevar a cabo la segmentación se deberán identificar las diferentes formas de dividir el mercado, desarrollando perfiles y descripciones de los segmentos resultantes.

A continuación encontramos algunas estadísticas que indican el potencial existente en las redes sociales para segmentar el mercado por clientes, productos o servicios:

- El 79% de los internautas utiliza redes sociales.
- El 51% de los usuarios de Facebook sigue a una marca comercial, en Twitter este porcentaje es del 30%.
- El 19% de los usuarios de Facebook hace comentarios acerca de los productos que ha comprado o le gustaría comprar.
- El 8% de los usuarios de Twitter se pondrán en contacto con el servicio de atención al cliente de una marca a través de este canal.
- El 11% de usuarios hablará con sus amigos de los productos y marcas que les gustan y el 15% comentará la publicidad de la empresa.

Gracias a los datos recopilados en redes sociales (“Me gusta”, número de seguidores, número de veces que se comparte un recurso, número de comentarios, edad, localización geográfica, etc.), las empresas tienen una fuente de información inmejorable para segmentar el mercado, crear nuevos productos que se adapten a las necesidades o gustos de los clientes, compararse y/o diferenciarse de la competencia y evaluar la forma más satisfactoria de llegar a los clientes.

1.3.2. Mejora de la atención al cliente

Las redes sociales permiten una interacción mucho más directa con los clientes. Antes de su existencia, las empresas realizaban la atención al cliente a través de canales como el correo ordinario, el teléfono o el email. La comunicación era mayoritariamente individualizada y muchas comunicaciones eran repetidas. Con las redes sociales, las empresas pueden comunicar con un número mayor de seguidores de forma más rápida y eficiente. Su mensaje es más directo y llega a más gente.

Por otro lado, los clientes aceptan que una empresa pueda tener una incidencia en algún momento, pero cuando esto ocurre, desean estar informados. En tal caso, los clientes querrán conocer el estado de la situación y su evolución. Por

ejemplo, supongamos que una empresa que ofrece servicios por Internet sufre una caída de sus servidores. Si la empresa gestiona su servicio de atención al cliente por teléfono, es muy probable que sus operadores se vean desbordados por el número de llamadas. Al no poder responder a tantas llamadas, muchos clientes recibirían la respuesta de “nuestras líneas están ocupadas, pruebe de contactar con nosotros más tarde” y el grado de satisfacción del cliente respecto a la empresa disminuirá. En cambio, si la empresa gestiona las incidencias mediante redes sociales, podrá informar a muchos más clientes, de forma proactiva y mucho más rápida, enviando mensajes de forma generalizada a todos sus clientes cada cierto período de tiempo hasta que la incidencia esté solucionada.

De esta manera, un problema para la empresa se puede convertir en una ventaja, ya que los clientes pueden ver que cuando hay un problema:

- La empresa comunica con el cliente de forma proactiva. Es la propia empresa la que toma la iniciativa.
- La comunicación con el cliente es constante y están siempre informados.
- En el caso de que exista un problema, es la propia empresa la que tiene cuidado de sus clientes y se ocupa de solucionar el problema en el menor tiempo posible.

1.3.3. Predicciones de mercado

Investigaciones recientes realizadas en las universidades de Manchester e Indiana han descubierto que Twitter es una herramienta útil para predecir el comportamiento del mercado bursátil. En el estudio se sostenía que el estado de ánimo y sentimental de las personas tienen una influencia directa sobre la tendencia del mercado y el precio de cotización de las acciones, de forma similar a la que puedan tener las noticias relacionadas con una empresa. También se ha demostrado que el número de tuits relacionados con una empresa es un buen indicador para predecir el aumento del precio de cotización de esa empresa al cierre del mercado bursátil en el mismo día.

En estos estudios se han alcanzado previsiones con un acierto del 87% en la predicción de la tendencia de las acciones, y los inversores que siguen tuits de los mercados de valores pueden incrementar sus ganancias en hasta un 15%.

Otro ejemplo del uso de Twitter en predicciones es el pronóstico, con un margen del error del 2%, del resultado final de las últimas elecciones alemanas.

Lectura recomendada

“¿Twitter puede predecir el mercado?”.
Finanzas.Personales.com.co
(2010).

Lectura recomendada

“Twitter predicts future of stocks”. *BBC News*
(06/04/2011).

1.3.4. Estudio de la competencia

Siempre se han podido estudiar las acciones de la competencia en mayor o menor medida. La principal ventaja que ofrecen las redes sociales es la facilidad en el acceso a los datos, su inmediatez y el reducido coste económico para su obtención.

Con las redes sociales podemos estimar lo que opinan los usuarios de la competencia de una empresa, sobre qué productos muestran más interés, dónde reciben quejas, sobre qué productos o servicios realizan acciones publicitarias, quiénes son los líderes de opinión, qué mensajes transmiten estos líderes, etc.

Cabe destacar que las empresas de la competencia pueden realizar las mismas acciones que nosotros, ya que gracias a las redes sociales, el conocimiento de nuestro mercado y nuestra situación en él se puede realizar de forma más fácil y económica.

1.3.5. Acceso a nichos de mercado

Las redes sociales permiten identificar y llegar a los nichos de mercado más fácilmente. Antes de la aparición de la web 2.0, estos nichos ya existían pero debido a la distancia geográfica de los integrantes del nicho, o a una mayor dificultad en la comunicación, eran altamente inaccesibles o permanecían ocultos.

Las personas pertenecientes a un nicho acostumbran a comunicarse mucho entre ellos y a buscar soluciones a sus problemas de forma activa. Pero aquellos integrantes del nicho que estén lejos geográficamente pueden tener dificultades de comunicación y por lo tanto, ver insatisfechas sus necesidades. A día de hoy, estas personas se pueden comunicar de forma fácil y rápida por las redes sociales. El nicho ha dejado de ser inaccesible para usuarios que no viven en grandes ciudades. Por otro lado, muchas empresas han podido desarrollar productos nuevos debido a la eliminación de las barreras geográficas y al hecho de que muchos comercios ya no son tiendas físicas, sino que ofrecen productos y servicios *online*.

Un nicho de mercado favorecido por la Web 2.0 ha sido el de los productos alimentarios sin gluten para celíacos. A día de hoy, podemos encontrar tiendas *online* específicas para celíacos, se han creado productos nuevos para ellos y gracias a la facilidad de comunicación, se han creado guías de productos alimentarios, hoteles y restaurantes.

1.3.6. Estudios de mercado

La gran cantidad de datos obtenidos de las redes sociales permite realizar estudios de mercado con mayor detalle, en menos tiempo y de forma más económica. La opinión de los clientes o usuarios puede ser recogida de forma

Lecturas recomendadas

“Cero TACC, celíacos conectados por medio de una red social”. Vitónica (19/12/2011).

Podéis consultar también la página web de Cero TACC.

automática en tiempo real, y la información obtenida no se limita solo a los usuarios de una muestra representativa del grupo a estudiar, sino que puede abarcar al conjunto de todos los usuarios que comentan en las redes sociales sobre el tema de interés.

Hasta hace unos años, la forma de medir la popularidad de un programa de televisión o de una persona se hacía mediante un estudio de medios, encuestas telefónicas o encuestas a pie de calle. En la actualidad, la opinión que emiten los usuarios sobre un personaje público o un programa puede hacer que el programa o persona aumente su popularidad o caiga en desgracia en un abrir y cerrar de ojos. Además, gran parte de estos comentarios son realizados en redes sociales simultáneamente a la visualización del programa, por lo que las redes sociales se han convertido en un medidor rápido, fiable y económico de los medios audiovisuales. Tanto es así, que la consultora Nielsen ha lanzado un medidor de audiencias televisivas en algunos países a partir de los comentarios en Twitter.

1.3.7. Comunicación inter e intraempresas

Varias plataformas y redes sociales han supuesto un ahorro importante en tiempo, dinero y facilidad de comunicación para las empresas. Además se pueden utilizar sistemas de BI de forma colaborativa, compartiendo cuadros de mando entre directivos geográficamente dispersos y de distintas empresas. Esto permite una visión global de la situación y una visión segmentada de su zona geográfica o segmento, fomentando la discusión y la toma de decisiones de forma colaborativa. Además, las empresas pueden realizar presentaciones y vídeollamadas *online* de forma simultánea en varios puntos del planeta, retransmitir en directo conferencias y, mediante wikis, editar simultáneamente varias personas un documento viendo todas ellas los cambios realizados en tiempo real.

1.4. Uso de datos móviles en sistemas de BI

Con la aparición de los teléfonos inteligentes, la cantidad de datos generada e intercambiada ha crecido exponencialmente.

El teléfono móvil une el mundo *offline* y el mundo *online*, permitiendo consultar cualquier contenido web allá donde estemos. Los nuevos dispositivos móviles nos ofrecen servicios de oficina virtual en un dispositivo de bolsillo, permitiendo realizar muchas tareas desde fuera de la oficina, suponiendo un ahorro de tiempo y dinero. Además, estos dispositivos permiten un acceso inmediato a los datos desde cualquier lugar: cuentas de clientes, *stock*, catálogo de productos, etc.

Sin duda, otra de las ventajas (o inconveniente, según se mire) de los teléfonos móviles es que permiten monitorizar la posición de los usuarios. Con toda esta información georreferenciada se pueden definir todavía mejor los perfiles

Referencia bibliográfica

"Nielsen launches 'Nielsen Twitter TV Ratings'" (10/07/2013).

Nielsen

Nielsen es una empresa consultora en marketing encargada, entre otros, de la medición de audiencias televisivas.

de los usuarios, sabiendo el tiempo que pasa en una tienda concreta, en un parque, cuándo toma una foto, qué recorrido hace y las paradas intermedias que realiza. Si hace una foto y la sube a las redes sociales, podemos saber el tiempo que pasa en el lugar donde toma la foto, si se encuentra delante de un monumento, podemos saber también el tiempo que dedica a observarlo, etc. La georreferenciación aún no ha llegado a sus máximas cotas, ya que hoy en día solo disponemos de forma masiva de información de exteriores (también denominada *outdoor*). El día en que también se disponga de información de interiores (también denominada *indoor*) las posibilidades de negocio van a crecer exponencialmente, ya que cada tienda física podría tener información de qué productos está mirando un usuario, durante cuánto tiempo, etc.

Por todo ello, podemos ver a los dispositivos móviles como al gran hermano que todo lo ve. Obviamente, el usuario de estos dispositivos no ofrecerá sus datos de forma indiscriminada, pero quizá se anima a hacerlo si a cambio de la pérdida de privacidad le proporcionamos conveniencia y practicidad, como, por ejemplo, una aplicación que lo guíe por el supermercado para indicar la ubicación de los productos de su lista de la compra, y que le facilite productos alternativos en caso de que alguno de los de su lista no esté disponible. Otros ejemplos serían aplicaciones turísticas donde el dispositivo móvil actúa de audioguía ofreciendo información de los elementos que el usuario tiene ante sí o una aplicación de viaje que ponga en contacto al usuario con personas que compartan sus intereses. Ejemplos de este último tipo de aplicación son Grindr, Twoo o MeetMe. Se trata de aplicaciones que permiten contactar con usuarios en una zona geográfica cercana que disponen de tiempo libre o con las que se tienen intereses en común. Son sistemas que permiten ver el perfil de los usuarios disponibles y de los eventos que se celebran en la zona.

El uso de información de dispositivos móviles en los procesos de inteligencia de negocio es lo que se conoce como *Mobile BI*. Este está intrínsecamente ligado y podría verse como un caso particular del BI social, con mucha más complejidad añadida. Entrar en detalle en el Mobile BI está fuera del objetivo de este módulo, pero hemos creído relevante referenciarlo aquí porque mucha de la información publicada en redes sociales proviene de dispositivos móviles. A lo largo del módulo, en algunos puntos se irá matizando o entrando en más detalle en algún aspecto que tenga que ver con datos móviles cuando se crea necesario.

2. Métricas e indicadores

Galileo Galilei dijo que “midiéramos todo lo que fuera medible e hiciéramos medible lo que no lo fuera”. Las redes sociales ofrecen una gran cantidad de métricas e indicadores para medir el progreso de los objetivos planteados. En este apartado estudiaremos métricas que nos permiten obtener una representación fiel de la imagen de nuestra empresa o negocio en las redes sociales, y los indicadores o KPI específicos de redes sociales que nos pueden ayudar a definir y medir el progreso en las redes sociales de los objetivos de la empresa.

2.1. Métricas en redes sociales

Las métricas que podemos obtener de las redes sociales son muchas y pueden variar en función de cada red social: número de “Me gusta” en Facebook, número de *followers* en Twitter, visualizaciones de vídeos en Youtube y un largo etc.

Debido a que los objetivos de una empresa o negocio varían con el tiempo, es posible que una misma métrica no pueda utilizarse siempre como indicador para medir los objetivos de la empresa, pero en cambio, un conjunto de métricas sí que puede reflejar fidedignamente la situación de nuestra empresa o negocio en las redes sociales.

De entre el conjunto de métricas disponibles en las redes sociales con las que trabajemos, deberemos escoger aquellas que den respuesta a los siguientes conceptos:

1) **Visibilidad y exposición:** Debemos conocer el número de personas que nos han visto *online*. Para ello habrá que controlar, entre otros:

- Cantidad de tráfico hacia el sitio web.
- Cantidad de visitas.
- Cantidad de las páginas vistas.
- Número de *followers*.
- Número de suscriptores.

2) **Sentimiento y notoriedad:** La notoriedad hace referencia al número de conversaciones sobre una marca frente a su competencia. Se deberán seguir todas las menciones de una marca y de su competencia en un periodo determinado de tiempo. Se deberá estudiar el sentimiento positivo, negativo y neutro de los usuarios, para conocer el sentimiento global hacia la marca.

3) Influencia: La influencia en redes sociales es la capacidad de un usuario de modificar el comportamiento o la toma de decisiones de sus seguidores. Un influenciador, también denominado líder de opinión, es una persona de la cual la audiencia se fía o confía, así que, cuando un influenciador comparte opiniones, pensamientos, reflexiones o ideas, la audiencia está más dispuesta a adoptar su mensaje o estar de acuerdo con él. Para medir la influencia de un contenido se puede utilizar:

- El número y la calidad de los *links* entrantes a un contenido.
- Los *links* de Twitter que son retuiteados o que se comentan.
- Comentarios y “Me gusta” de los *posts* en Facebook.
- Contenido que es compartido y recomendado (y en qué medida).

4) Engagement o vinculación emocional, es decir, la medida en la cual la gente interactúa con una empresa (marca o usuario) y su contenido. El *engagement* muestra las personas que están interesadas en lo que la empresa ofrece y están dispuestas a participar. Medir el *engagement* es importante para ver cuánto y cómo de frecuente el público participa e interactúa con el contenido de la empresa. Para medir el *engagement* se puede utilizar:

- Los nuevos “Me gusta” de la página de la empresa.
- Número de veces que se ha compartido alguna información proporcionada por la empresa.
- Menciones (positivas, negativas o neutras).
- Comentarios en el *blog*.
- Valoraciones.
- Visitas de imágenes/vídeo.

5) Popularidad: La popularidad *online* es básicamente el número de gente que se suscribe al contenido de una empresa (marca o usuario). Para medirlo podemos utilizar:

- Número de suscriptores al RSS o al *email*.
- Número de *followers* en Twitter.
- Número de miembros en tu grupo de LinkedIn.
- Número de “Me gusta” en la página global de Facebook, no en un contenido en concreto.

2.2. Indicadores

Un indicador o KPI es una métrica que ayuda a definir y medir de forma clara el progreso hacia los objetivos de empresa y que, a su vez, es de fácil trazabilidad. En función del objetivo perseguido por la empresa, deberemos centrarnos en aquellos KPI que mejor ayuden a conseguir los objetivos de la empresa, intentando reducir el número de KPI necesarios. Recordemos que:

Recordatorio

KPI son las siglas de Indicador clave de desempeño (*Key performance indicator* en inglés) y se utilizan para medir el nivel de desempeño de procesos para alcanzar objetivos fijados.

“10 KPI son buenos, 5 KPI son mejores y 1 KPI es ideal.”

Los KPI deberán estar bien definidos y ser realistas. Al diseñarlos, es conveniente seguir la filosofía SMART, cuyas siglas indican cómo deben ser los KPI de un sistema:

- ESpecíficos (*Specific*).
- Medibles (*Measurable*).
- Alcanzables (*Achievable*).
- Relevantes (*Relevant*).
- A Tiempo (*Timely*).

Cuando trabajamos en redes sociales, al definir los KPI significativos de una red social, habrá que diferenciar claramente los **elementos principales** que la forman, que normalmente podremos clasificar como recursos y agentes (también denominados usuarios). A continuación mostramos algunos ejemplos:

- En Facebook diferenciaremos los **recursos** (fotos, mensajes, invitaciones, vídeos, enlaces, etc.) de las **personas** o **empresas** (usuarios de la red social que interactúan con los recursos).
- En Twitter los recursos serán los **tuits** (mensajes que los usuarios escriben) y los **hashtags** (palabras clave que identifican una temática o a una serie de mensajes). Los usuarios (personas o empresas) son fácilmente identificables, ya que su identificador va precedido de una arroba.
- En Youtube tendremos **usuarios, vídeos y canales**.
- En Pinterest o Flickr tendremos **usuarios, fotos y etiquetas**.

Una vez hayamos diferenciado los recursos y los agentes de la red social, podremos definir los KPI. A continuación enumeramos algunos ejemplos de KPI para Facebook, Twitter y Youtube:

1) En Facebook:

- Total de “Me gusta” en página.
- Nuevos “Me gusta” en página (diarios, semanales, mensuales).
- Segmentación de los “Me gusta” recibidos según varios parámetros: por contenido relativo a una sección de las páginas o departamentos de la empresa, por tipo de contenido creado por la empresa o que trata sobre ella (entrada explicativa sobre un hecho o evento realizado por la empresa, explicación o aportación por parte de la empresa sobre una noticia publi-

cada por un tercero, sobre los comentarios que han realizado los clientes de la empresa), etc.

- Fuente de nuevos “Me gusta” en página:
 - Páginas sugeridas.
 - En página.
 - Móvil.
 - A través de publicaciones.
 - Otras fuentes.
- Personas hablando del tema.
- Mensajes en el muro.
- Engagement.
- “Me gusta” en publicaciones.
- Comentarios en publicaciones.
- Compartir en publicaciones.
- Total de visitas en página de fans.
- Procedencia de visitas.
- Alcance de publicaciones.
- Clics en publicaciones.
- Clics en enlaces.

2) En Twitter:

- Total seguidores.
- Nuevos seguidores de perfil (diarios, semanales, mensuales).
- *Engagement*.
- Menciones.
- Respuestas.
- Retuits.

- Favoritos.
- Impresiones.
- Clics en enlaces.
- Listas en las que una cuenta aparece.

3) En Youtube:

- Total de reproducciones por canal.
- Nuevas reproducciones en el canal durante la última semana.
- Total de suscriptores.
- Nuevos suscriptores.
- *Engagement*.
- “Me gusta”.
- “No me gusta”.
- Comentarios.
- Vídeos compartidos.
- Vídeos favoritos.
- Favoritos eliminados.
- Fuente de tráfico.
- Aplicaciones para móviles y tráfico directo.
- Referencias de reproducciones de Youtube.
- Referencias de reproducciones externas a Youtube.

3. Herramientas de medición

Las herramientas que se utilizan en las soluciones de BI deben ser útiles para la toma de decisiones y deben mostrar la información de forma resumida y en tiempo real, o en un breve lapso de tiempo. En el mercado hay gran cantidad de herramientas que permiten analizar los datos de las redes sociales, desde soluciones a medida, hasta soluciones globales como Google Analytics. Algunas de estas herramientas también permiten procesar información extra proveniente de dispositivos móviles. En este apartado empezaremos tratando las herramientas de BI Social clásicas, para luego entrar en más detalle en las que se han especializado en el análisis de información proveniente de dispositivos móviles.

3.1. Herramientas de medición de datos de redes sociales

A continuación clasificamos las distintas herramientas para el análisis de información de las redes sociales en función del objetivo del análisis a realizar:

1) **Herramientas de análisis:** Ofrecen información sobre el tráfico procedente de redes sociales y permiten obtener estadísticas del tráfico segmentando por campaña y red social. Normalmente permiten crear nuevos indicadores y cuadros de mando para personalizar las estadísticas realizadas y la información proporcionada. Como ejemplos de herramientas de análisis encontramos Google Analytics y Adobe Marketing Cloud.

2) **Herramientas de monitorización:** Permiten hacer un seguimiento de lo que “se habla” en las redes sociales respecto a una marca, un producto o una campaña. Permite también monitorizar los usuarios que generan los comentarios. La monitorización se realiza a partir de un conjunto de palabras clave y de canales escogidos con antelación. Es sumamente importante definir un conjunto de palabras relevantes para realizar la monitorización y la consiguiente toma de decisiones correctas. Entre las herramientas de monitorización disponibles destacan Radian6, Brandwatch, Talkwalker, SocialMention, Howsociable y Addictomatic.

3) **Herramientas de reporting:** Muestran estadísticas sobre campañas *online* y a menudo ofrecen también opciones de análisis. Muchos canales de social media ofrecen herramientas de *reporting* avanzadas como Facebook Insights, Twitter Analytics, y Youtube Analytics. No obstante, en algunos canales, como por ejemplo Google Plus, LinkedIn y Pinterest, las herramientas aún ofrecen estadísticas muy básicas. Además de las proporcionadas por los propios canales, existen también otras herramientas, como son Crowdbooster, TrueSocial-Metris y Cyfe.

4) Herramientas de *benchmarking*: Permiten comparar la evolución de una empresa (producto o campaña) en las redes con la evolución de los competidores. Estas herramientas se pueden configurar para que envíen alertas cuando las páginas de la competencia realicen cambios o cuando empiecen a ofrecer un nuevo servicio.

Entre las herramientas disponibles destacan SocialBakers, DKS Social Smart, Twittercounter, chengedetecion.com, Compete Monkey y Wildfire Social Media Monitor.

5) Herramientas de *management*: En general, se trata de herramientas multifunción para el día a día y de muy fácil manejo. Se utilizan para publicar, programar y gestionar la presencia social de una empresa (producto o campaña) en un gran número de cuentas de redes sociales. Pero también pueden ofrecer análisis iniciales, comparativas con competidores e informes. Entre estas destacan Tweetdeck, Hootsuite, Sprout Social y Buffer.

Existen muchas otras herramientas para la medición en redes sociales. En cada proyecto será necesario decidir qué herramientas utilizar y si hay que integrar todos los aspectos del negocio en una sola herramienta o, por el contrario, es conveniente utilizar diferentes herramientas para medir los resultados en diferentes medios. A continuación, y a modo de ejemplo, mostramos cómo funciona una de estas herramientas: Facebook Insights.

3.2. Ejemplo: Facebook Insights

Facebook Insights es una de las herramientas más completas para medir la actividad de Facebook. Su página principal, mostrada en la figura 2, enseña un resumen de la actividad de la última semana relacionada con un usuario de Facebook. En particular, muestra los “Me gusta” recibidos en la página, el alcance de las publicaciones, la participación de los usuarios con la página y las cinco publicaciones más recientes.

Web recomendada

La red latinoamericana de portales educativos ofrece un listado de 70 herramientas de medición en redes sociales clasificadas por canal u objetivo <http://www.relpe.org/ultimasnoticias/70-herramientas-para-monitorear-las-redes-sociales/>

Figura 2. Página principal de Facebook Insights



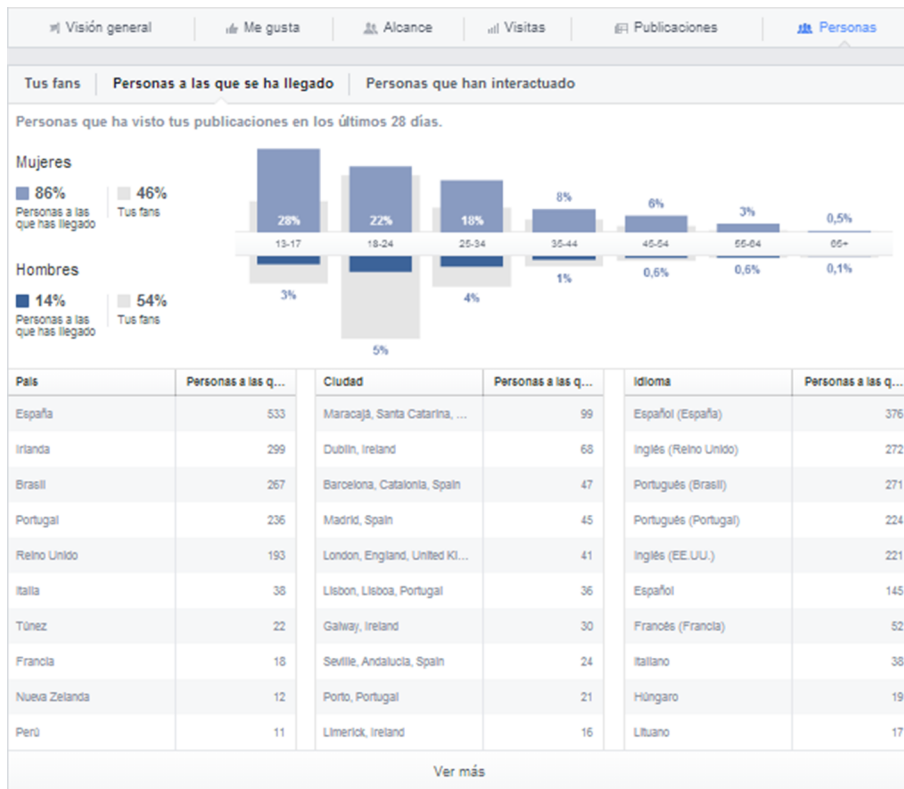
Desde el menú superior se puede acceder a diferentes secciones que detallan los elementos más importantes de la actividad de una página en Facebook:

- La página de “Me gusta” contiene la información relativa a los “Me gusta” recibidos en la página. En particular, ofrece información detallada sobre:
 - La evolución cronológica del total de “Me gusta”.
 - La evolución diaria de los “Me gusta” recibida a través de publicidad pagada en Facebook durante el último mes.
 - La evolución diaria de los “Me gusta” orgánicos¹ el último mes.
 - La evolución diaria de los “Ya no me gusta” del último mes.
 - El origen de los “Me gusta” en función del origen de ellos (dispositivos móviles, sugerencias de páginas, publicaciones de otras personas, etc.).
- La página de “Alcance” muestra información sobre el número de personas que han visto las publicaciones de la página. Permite diferenciar entre el alcance pagado a través de publicidad que hacemos en Facebook y el orgánico, y entre el alcance proveniente de los “Me gusta”, de los comentarios y de las veces que se ha compartido un contenido de la página analizada.
- La página de “Visitas” muestra la información sobre el número de veces que se ha visto cada una de las secciones de la página (biografía, pestaña de información, pestaña de fotos, etc.) y sobre las visitas que han llegado desde fuera de Facebook.
- La página de “Publicaciones” indica las horas a las que se conectan los fans de la página durante la semana. También es posible visualizar esta información diariamente y por cada publicación.

⁽¹⁾ Los clics orgánicos son aquellos que se han recibido de forma gratuita a partir del contenido generado por nuestra página, por el contenido compartido o por los comentarios escritos en nuestros artículos.

- La página de “Personas” muestra las personas a las que les gusta la página analizada, el número de personas a las que se les mostró información sobre la página en su muro (ya sea una publicación o un comentario propio de la página, o un comentario “Me gusta” realizado por algún amigo suyo) y personas que han interactuado con la página web o su contenido. La información se segmenta por sexo, franjas de edad, país de procedencia e idioma, como puede verse en la figura 3.

Figura 3. Segmentación de los visitantes de la página de Facebook



3.3. Herramientas de medición de dispositivos móviles

Normalmente, medir datos procedentes de dispositivos móviles tiende a ser más difícil que en redes sociales porque las acciones realizadas con dispositivos móviles son mucho más diversas y están más contextualizadas.

Algunas herramientas denominadas de análisis, que se anuncian como herramientas de Mobile BI, se limitan a adaptar su apariencia a un formato *mobile-friendly*, sin introducir ningún cambio significativo para tratar datos provenientes de móviles. Afortunadamente, hay otras herramientas que permiten analizar también datos obtenidos a través de aplicaciones móviles. Estas herramientas ofrecen funcionalidades bastante similares entre sí. Algunas serían:

- La carga personalizada de datos provenientes de dispositivos móviles en la herramienta de medición a través de formatos estándares como hojas de cálculo, ficheros XML o API.

- La generación de paneles de control personalizados.
- Una gran variedad de representaciones gráficas a mostrar en cada panel de control personalizado.
- Representación geoposicionada de los datos que contengan información sobre las coordenadas geográficas de donde tuvieron lugar.

Debido a la creciente importancia de las aplicaciones móviles, cada vez existen más herramientas para el análisis de datos provenientes de dispositivos móviles, y entre ellas hay que destacar a:

- Yellowfin Business Intelligence.
- Transpara.
- InetSoft.
- CollabMobile.
- JReport.

La característica principal que diferencia a estas herramientas de las herramientas de analítica web, o de medición de la actividad en redes sociales, es que permiten una carga personalizada de los datos y la generación de informes adaptados a cada necesidad.

La desventaja principal respecto a otras herramientas de medición es que los datos no son recuperados por la propia herramienta, como sí que pasa en Google Analytics o en Facebook Insights. A pesar de que las herramientas aceptan varios formatos para cargar los datos en la aplicación, el proceso de recuperación de los mismos se debe realizar aparte.

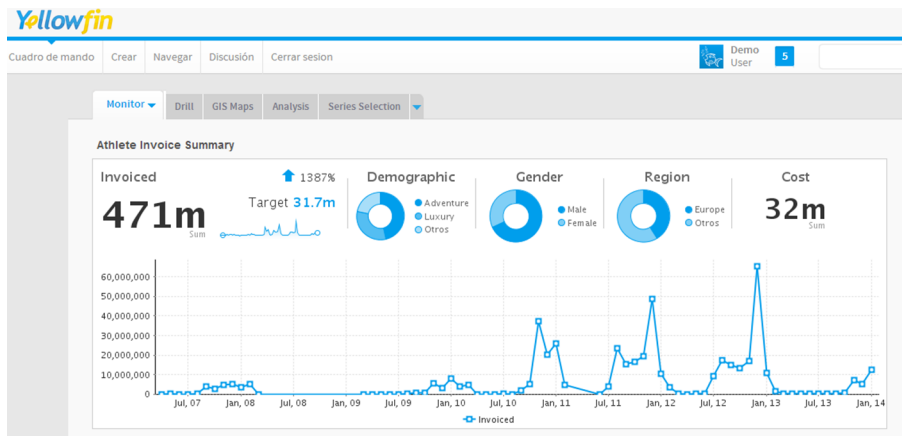
Su principal ventaja es la flexibilidad en su personalización, que permite representar distinta información en cada empresa, y la gran variedad de cuadros de mando que son capaces de generar.

A continuación, y a modo de ejemplo, mostramos cómo funciona una de estas herramientas: Yellowfin Business Intelligence.

3.4. Ejemplo: Yellowfin Business Intelligence

Yellowfin Business Intelligence es una de las herramientas de medición más completas para medir datos provenientes de dispositivos móviles. Su página principal, mostrada en la figura 4, enseña la pestaña del panel de control que hayamos definido como página principal del cuadro de mando.

Figura 4. Página principal de Yellowfin Business Intelligence



En cada página de la herramienta tenemos acceso a dos menús. El menú inferior nos permite acceder a cada una de las secciones que forman el cuadro de mando.

Por su parte, desde el menú superior, se puede acceder a diferentes secciones que detallan los elementos más importantes para la configuración y utilización de Yellowfin:

1) **La pestaña "Cuadro de mando".** El cuadro de mando puede contener tantas secciones como creamos oportunas y existen muchos tipos de representaciones gráficas diferentes que podemos utilizar (histogramas, gráficos de barras, de líneas, de áreas, etc.), así como tablas de datos y representaciones geolocalizadas en un mapa. Tanto las pestañas que contendrá el cuadro de mando, como el tipo de información que se mostrará en cada una de las pestañas, deberemos definirlos nosotros a partir de la opción de "Crear" del menú superior. Las figuras 5 y 6 muestran dos cuadros de mando de Yellowfin que permiten mostrar los datos mediante gráficos tradicionales y mediante mapas.

Figura 5. Ejemplos de representaciones gráficas

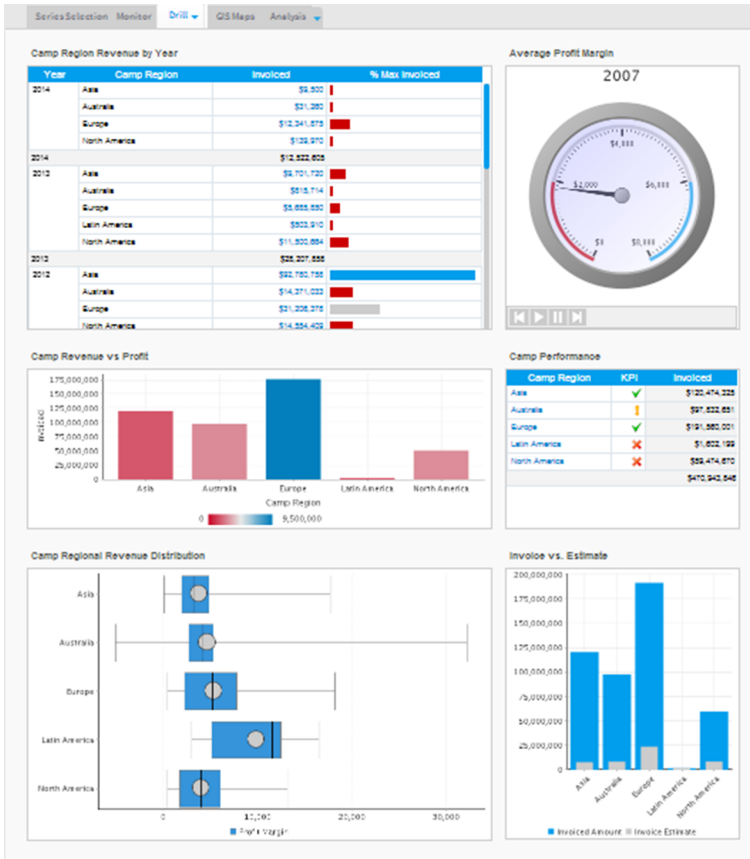
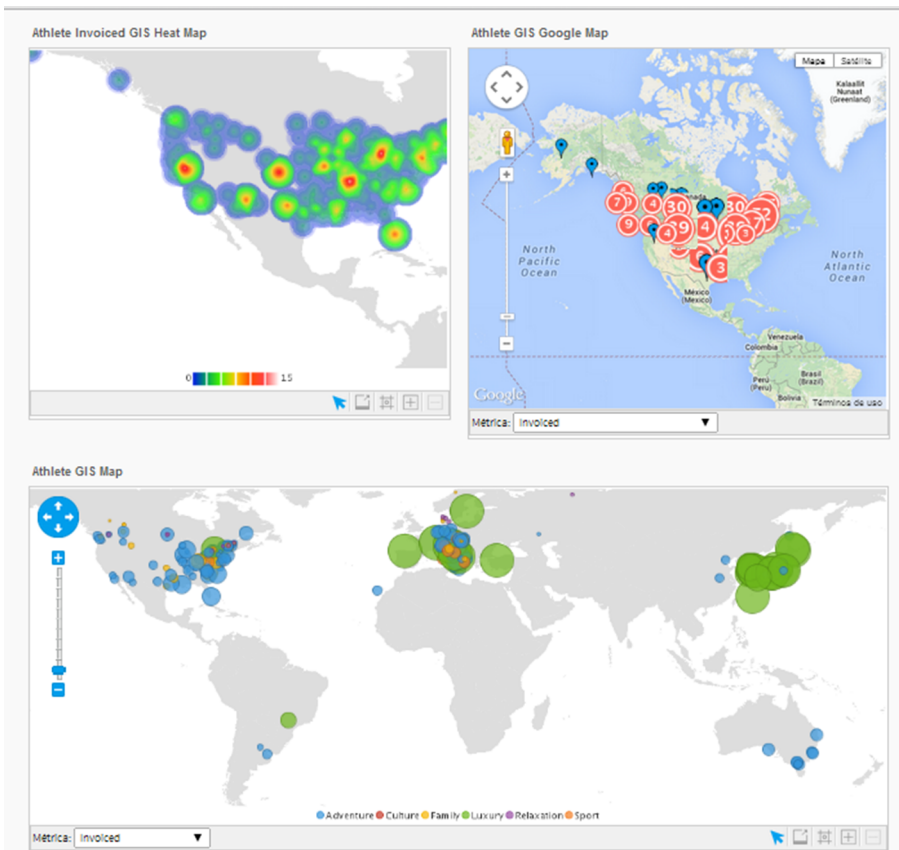


Figura 6. Ejemplos de representaciones geopositionadas



2) La pestaña “Crear”. Nos permite escoger entre crear un tipo de informe para el cuadro de mando, o bien crear una nueva página para el cuadro de mando que contenga alguno de los tipos de informes ya creados.

3) La pestaña “Creación de informes”. En la pestaña de “Creación de informes”, tendremos que seleccionar las fuentes de datos que queramos utilizar para elaborar el informe, los filtros a aplicar en algunos atributos de los datos, el tipo de representación a utilizar (gráfica, tabla, georreferenciada, etc.) y la fórmula que se utilizará para calcular la representación.

4) Pestaña “Creación de páginas del cuadro de mando”. En la pestaña de creación de páginas del cuadro de mando, se indicará la posición del cuadro de mando que deberá ocupar la nueva pestaña y el conjunto de informes o representaciones gráficas que deberá contener esta página.

5) Pestaña “Navegar”. En la pestaña “Navegar”, tenemos acceso de forma estructurada al conjunto de datos cargados en la herramienta, las páginas definidas en el cuadro de mando, los KPI utilizados, y un tutorial detallado sobre el funcionamiento de la herramienta.

Como se puede observar en la figura 7, los datos mostrados en la pestaña “Navegar” están organizados en carpetas. Si seleccionamos un documento dentro de una carpeta, se mostrará una explicación detallada del tipo de información que contiene la opción seleccionada, que podrá dar una descripción del cuadro de la opción de la métrica o pestaña, o del cuadro de mando, una explicación del KPI, los datos cargados (si se trata de un conjunto de datos) y una presentación gráfica (si se trata de un tutorial sobre Yellowfin).

Figura 7. Detalle de la pestaña “Navegar”

The screenshot shows the Yellowfin interface with the 'Navegar' (Navigate) tab selected. The top navigation bar includes 'Cuadro de mando', 'Crear', 'Navegar', 'Discusión', and 'Cerrar sesión'. The user is identified as 'Demo User'.

The main content area is divided into three sections:

- Buscar (Search):** Includes a search bar and a 'Buscador Avanzada' (Advanced Search) button.
- Marketing & Booking:** A sidebar menu with options like 'Reportes explorando el proceso de booking y campañas de marketing', 'Informes: 9', and 'Historial de las versiones salvadas'.
- Contenido (Content):** A table listing various reports and tutorials. The 'Marketing & Booking' folder is expanded, showing items like 'Agency Benchmark', 'Agency Sales over Time', 'Athlete Booking by Method', 'Booking Process Donut', 'Cancelled Bookings', 'Invoice vs. Estimate', 'Marketing Analysis', 'Media Category Profit', 'Top N Agencies', and 'Training'.
- MI Contenido (Content):** A list of 'Mis favoritos' (My favorites) and 'Acceso reciente' (Recent access) items, including 'Athlete', 'Camp Region Revenue', 'Agency Benchmark', 'Athlete Cost Summary', 'Camp Revenue vs Profit', 'Athlete Profit Summary', 'Agency Sales over Time', 'Athlete Demographic', 'Camp Monthly Perform...', 'Camp Revenue Detail R...', and 'Profit'.

4. Ejemplo de BI en redes sociales

Al igual que en los proyectos de BI tradicionales, en el BI social es necesario saber qué preguntas se pretenden responder. Por lo tanto, una estrategia de social media debería definir claramente los objetivos, evaluar las oportunidades y seleccionar el medio de comunicación apropiado. A continuación mostramos un caso de éxito para ejemplificar cómo realizar esta tarea.

Tomamos como ejemplo Hokey Pokey Ice Cream Creations, una empresa de la India encargada de la venta de helados *premium* con distribución en una docena de tiendas propias. Mediante campañas en redes sociales, la compañía quería incrementar el conocimiento de su marca, incrementar sus ventas y aumentar y fidelizar clientes. Con la colaboración de la universidad de Harvard crearon una estrategia en un periodo de 30 meses basada en los siguientes 7 pasos:

1) Recuperar las conversaciones de Facebook y Twitter sobre la marca:

Se recuperan los mensajes relacionados con la marca escritos por usuarios de Facebook y Twitter, permitiendo evaluar la percepción que tiene el público sobre la empresa y acceder a personas identificadas con la marca.

2) Identificar líderes de opinión: El objetivo era identificar, a partir de los mensajes analizados en el paso anterior, posibles líderes de opinión (respetados y seguidos en la red), que pudieran convertirse en embajadores de la marca para transmitir la campaña publicitaria creada por la empresa para las redes sociales.

3) Evaluar a los líderes de opinión: El perfil de cada posible líder de opinión seleccionado se evaluó en función de un conjunto de parámetros para medir su influencia en la red:

- Propagación de mensajes: cuántas veces se propagan (valor medio) sus mensajes sobre temas relacionados con Hokey-Pokey.
- Influencia: número de conexiones y/o *followers* que tiene.
- Locuacidad: número de usuarios que han propagado algún mensaje suyo.
- Impacto social: número de comentarios y respuestas obtenidos por cada mensaje sobre temas relacionados con Hokey-Pokey.
- Similitud: similitudes y temas de interés comunes entre los posibles líderes y sus redes de contactos.

4) Identificar a potenciales embajadores: Para identificar a posibles embajadores se identificó a los líderes de opinión resultantes del paso anterior que tenían interés sobre los temas o categorías a las que pertenece el producto, en este caso, helados. Un buen embajador de helados es alguien con reputación en este ámbito.

5) Reclutar a los embajadores: Se contactó con los embajadores seleccionados para que formaran parte de la campaña creada por la empresa.

6) Ejecución de la campaña: Durante la campaña, los embajadores transmitieron el mensaje de la empresa en la red social. Eso se hizo en dos etapas:

- Creaciones en el muro: En esta etapa los embajadores tenían que crear sus propios helados personalizados, darles un nombre, explicar el proceso de creación y publicar sus recetas en la web de Honey-Pokey. En dicha web también se mostraban los productos de la marca, sus tiendas y las nuevas creaciones.
- Campañas virales: En esta etapa los embajadores compartían sus creaciones con sus seguidores en Facebook y Twitter. El mensaje transmitido tenía que identificar la creación con el embajador que la creó e incentivar la compra mediante descuentos a sus seguidores.

7) Medir los resultados de la campaña: En la evaluación de la campaña se relacionaron métricas de social media, como conversaciones y mensajes con métricas financieras para calcular el incremento de presencia de la marca y sus ganancias.

Los resultados de la campaña fueron satisfactorios. El conocimiento de la marca Hokey-Pokey se incrementó en un 49%. Los ingresos crecieron un 40% y el ROI fue del 83%. El 80% de los ingresos de la campaña viral fue atribuible a Facebook, el 23% a Twitter y hubo entre un 3% y 8% de solapamiento entre las dos redes sociales.

5. Almacenamiento y recuperación de datos de redes sociales

Sin lugar a dudas, uno de los puntos que genera más trabajo e investigación en el Social Media BI y en el BI Mobile es su gran capacidad para generar datos. Esos datos son de vital importancia para los profesionales y empresas, que ven cómo, a partir de ellos, pueden extraer información útil para dar soporte a la toma de decisiones. En este apartado vamos a introducir ambos aspectos, explicando cómo extraer los datos de redes sociales y qué debemos tener en cuenta al almacenar estos datos en bases de datos.

5.1. Recuperación de información de redes sociales

Las redes sociales permiten recuperar parte de su información con el uso de *plugins* y *widgets* o bien a través de consultas a su API.

Los *plugins* y *widgets* para redes sociales son pequeñas aplicaciones o programas que permiten visualizar parte del contenido de la red social y realizar algunas acciones sobre ella, como comentar artículos o seguir a un usuario desde una web o aplicación móvil externa a la red social.

Los *widgets* y *plugins* están orientados al desarrollo de aplicaciones web y móviles, permitiendo una integración parcial de la red social en estas, pero no permiten la extracción de información de la red social para su posterior análisis.

Las API para redes sociales son bibliotecas informáticas que permiten a un programa externo, a través de consultas a la API, extraer datos de la red social para su posterior análisis y generación de información. Estas consultas se pueden personalizar con multitud de parámetros, como el idioma de los mensajes, la fecha y región de publicación, la persona que lo escribió, si pertenecen a una categoría, si contienen una etiqueta, etc.

Es habitual en las API la existencia de un límite máximo de consultas que una aplicación puede realizar durante un intervalo de tiempo. Cuando se intenta sobrepasar este límite, la API bloquea temporal o permanentemente la aplicación. De manera similar, también se limita el número de resultados devueltos en cada consulta, retornando a la aplicación únicamente los últimos *X* resultados correspondientes a la consulta realizada. El número de resultados varía mucho en función de la red social, y los valores pueden oscilar entre un máximo de 100 mensajes en Twitter a 5000 en Foursquare.

Las API son muy dinámicas y están en constante evolución. Son habituales las modificaciones sobre los límites de consulta, los parámetros para personalizar las búsquedas e incluso la forma de cómo deben ser consultadas, por lo que es necesario revisarlas muy a menudo.

A continuación describimos los *plugins* y las API de las principales redes sociales.

5.1.1. *Plugins* de las redes sociales

Los *plugins* de Facebook permiten hacer “Me gusta” en un contenido externo de la red social, visualizar las personas a las que les gusta una página de Facebook, hacer comentarios en una página externa y que estos se vean tanto en la página original como en Facebook, mostrar en una página externa la actividad más reciente en el Facebook de un determinado usuario, compartir un contenido externo en un perfil de Facebook, etc.

Los *plugins* de Twitter permiten incrustar en una web externa el *timeline* de todos los mensajes de un usuario, lista o búsqueda, y los botones para compartir una web externa en un tuit o para hacerse seguidor de un usuario de Twitter.

Los *widgets* de LinkedIn permiten visualizar el perfil de un usuario o empresa fuera de su web, y añadir los botones para hacerse seguidor de una empresa o persona y para compartir contenido dentro de tu perfil de LinkedIn.

5.1.2. API de las redes sociales

Respecto a las API, la información que proporcionan depende de cada caso. A día de hoy, quizás la más extensa de todas es la de Facebook. Esta permite:

- Crear una aplicación de uso interno en Facebook que permita desarrollar juegos *online*², relacionar usuarios, o simplemente promocionar un producto.
- Interactuar con una aplicación móvil para Android o iOS (iPhone) para añadir funcionalidades sociales a la misma, como por ejemplo compartir fotos, vídeos o publicar mensajes en el muro de los usuarios o en el de sus amigos.
- Realizar búsquedas de elementos relacionados en Facebook mediante el Open Graph Search. El API del Open Graph Search permite buscar contenido en Facebook de usuarios con los que no tenemos ninguna relación. Por ejemplo, podríamos recuperar información sobre personas que viven cerca de Barcelona a las que les gusta Queen, o personas que han asistido al último concierto de Lady Gaga en Madrid.

⁽²⁾Creación de juegos en Facebook: <https://developers.facebook.com/docs/games> y <https://developers.facebook.com/docs/games/canvas>.

- Mediante la API del Open Graph Search podemos realizar consultas de cualquier relación entre usuarios de Facebook, o entre usuarios y elementos de Facebook (fotos, vídeos, eventos, “Me gusta”, etc.) que las políticas de privacidad seleccionada por un usuario permita.

La API de Twitter permite crear aplicaciones con las que seleccionar de forma automática información de los usuarios de Twitter, de sus seguidores o de los usuarios a los que sigue, de los tuits que ha publicado, etc. En las búsquedas se puede:

- Definir el idioma de la búsqueda y el de los resultados.
- Geolocalizar la búsqueda a un país o región utilizando coordenadas geográficas y un radio en kilómetros o millas, permitiendo seleccionar solamente mensajes emitidos en esa zona.
- Indicar el intervalo de tiempo de publicación de los mensajes.
- Seleccionar el tipo de tuits, diferenciando entre recientes, populares o una combinación de ambos.

La API de LinkedIn permite realizar búsquedas sobre empresas, trabajos, personas, grupos de personas, mensajes publicados en los perfiles de estos elementos y las relaciones que existen entre ellos.

La API de Foursquare es una de las que ofrece más opciones de búsqueda. Permite realizar búsquedas geolocalizadas y sobre eventos que están teniendo lugar en estos instantes. También permite filtrar la búsqueda por categorías (restaurantes, compras, arte, aire libre, etc.) y por etiquetas. Permite ordenar o filtrar los resultados obtenidos a partir de las puntuaciones de los usuarios.

Para finalizar esta sección, se quiere remarcar que, en ocasiones, a los datos recuperados de las API de las redes sociales se les deberán aplicar filtros para extraer información de ellos, especialmente cuando trabajemos con cadenas de textos. En este caso podrá ser necesario recuperar solo las entradas que contengan una determinada palabra y/o aplicar técnicas para seleccionar también palabras escritas incorrectamente (con faltas de ortografía).

5.2. Bases de datos para datos procedentes de redes sociales

Tanto los datos provenientes de dispositivos móviles como de redes sociales tienen las principales características asociadas a los sistemas de datos masivos. Tienen un gran volumen de datos, son muy heterogéneos y su análisis debe hacerse de forma rápida porque un día en una red social puede significar

una eternidad. Y aún más, en un sistema móvil, un usuario necesita respuestas cuando está en un determinado lugar, no 10 minutos después de haberlo abandonado.

El hecho de tener datos tan voluminosos, variados y que requieren un tiempo de respuesta rápido, hace que nos debamos replantear el uso de sistemas gestores de bases de datos relacionales para almacenar y explotar los datos. Eso no quiere decir que no puedan utilizarse estos sistemas, sino que en determinados casos puede ser más eficiente usar otro tipo de base de datos.

Actualmente, hay gran diversidad de sistemas gestores de bases de datos. En particular, las bases de datos NoSQL pueden ser una muy buena alternativa para almacenar y procesar datos provenientes de redes sociales. A continuación enumeramos distintos tipos de bases de datos disponibles, junto con algunos casos de utilización en las redes sociales:

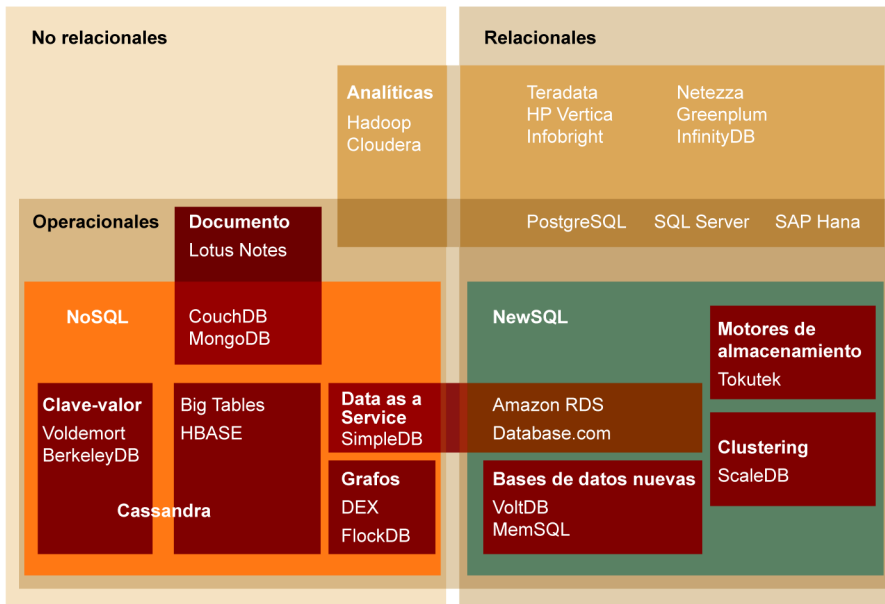
- **Bases de datos analíticas:** Permiten reducir el tiempo necesario para resolver determinado tipo de consultas. Por ejemplo, Yahoo utiliza Hadoop para calcular el precio de sus anuncios.
- **Bases de datos de clave/valor:** Permiten almacenar datos en formato clave-valor, donde la clave es un valor que identifica cada registro y los valores son vistos como cajas negras por la base de datos. En este tipo de bases de datos las consultas se realizan por la clave. Instagram utiliza Redis para guardar las sesiones de usuario y para almacenar los usuarios que han subido cada foto.
- **Bases de datos de documentos:** Permiten almacenar datos de forma agregada en documentos. Esto facilita la consulta de los datos en el caso de que la aplicación utilice el formato de agregado usado para la base de datos. Por ejemplo, Foursquare utiliza MongoDB para la información georreferenciada.
- **Bases de datos orientadas a columnas:** Organizan sus datos en filas y columnas en lugar de en filas como se hace en el resto de bases de datos agregadas (clave-valor y documental). Son idóneas para almacenar grandes cantidades de datos. Twitter utiliza Cassandra para su aplicación web.
- **Bases de datos en grafo:** Almacenan los datos en formato grafo y son idóneas para representar datos altamente relacionados. Google utiliza Pregel para relacionar páginas web.

Pero los sistemas gestores de bases de datos disponibles no se limitan a los mencionados. En la figura 8 podéis ver un esquema de las tecnologías más utilizadas actualmente para almacenar datos para soluciones BI.

Grafo

Un grafo es un conjunto de objetos llamados vértices o nodos, unidos por enlaces llamados vértices o aristas, que permiten representar relaciones binarias entre elementos de un conjunto.

Figura 8. Tecnologías de almacenamiento utilizadas en proyectos 2.0



Fuente: www.datknoisy.com.

6. Técnicas y algoritmos para la clasificación y personalización en social BI

Anteriormente hemos visto indicadores y herramientas que nos permiten analizar los datos de las redes sociales y dispositivos móviles. Sin embargo, hay muchos casos donde estos indicadores y herramientas no podrán dar una respuesta a nuestras necesidades, y tendremos que recuperar y analizar la información utilizando otras técnicas.

Debido a la gran cantidad de datos a analizar, a la necesidad de poder evaluar rápidamente nuevos casos cuando estos se produzcan y a la representación en forma de grafo de las relaciones entre individuos y entidades, las técnicas Data Mining y de teoría de grafos son las más apropiadas para trabajar con datos provenientes de redes sociales y de dispositivos móviles.

A continuación introduciremos algunos de los problemas más estudiados en Data Mining y que son de uso común en el análisis de datos de redes sociales: la agrupación y la clasificación de individuos. Posteriormente se describirán las técnicas de teoría de grafos que pueden ser de utilidad para analizar datos de redes sociales.

6.1. Agrupaciones de individuos

Cuando se tiene un gran número de seguidores o personas a las que les gusta una página, puede ser interesante clasificar a estas personas en grupos. Eso permitirá, por ejemplo, ofrecer a cada grupo solo aquella información que realmente sea relevante para ellos.

Los criterios para clasificar a un usuario en un grupo u otro pueden ser muy diversos, por ejemplo, se pueden analizar los tipos de recursos para los que el usuario indica “Me gusta” (películas de misterio, libros de humor, fotos de cantantes, etc.) o el tipo de personas a las que sigue. No obstante, independientemente de los criterios utilizados para agrupar a los individuos, los grupos deben satisfacer dos propiedades:

- Los individuos que componen cada grupo serán lo más próximos entre ellos. Distancia intragrupos pequeña.
- Los individuos de grupos diferentes serán lo más diferenciados posibles entre ellos. Distancia intergrupos grande.

Para la creación de grupos, también conocidos como *clusters*, se acostumbra a utilizar técnicas de **aprendizaje no supervisado** debido a que:

Ved también

Más información sobre las técnicas de agrupación y clasificación utilizadas en los sistemas de analítica de negocio se puede encontrar en los materiales de *Business Analytics*, de Jordi Girones Roig.

Aprendizaje supervisado y no supervisado

En el aprendizaje supervisado se parte de un conjunto de datos de los que se tienen las características de los usuarios y la clasificación para cada individuo. A partir de estos datos, el sistema puede aprender y ser capaz de predecir una solución más o menos acertada ante nuevos casos. En cambio, en el aprendizaje no supervisado, se tienen las características de cada usuario, pero no la solución del problema. Las técnicas de aprendizaje no supervisado son útiles para detectar grupos con características similares y permitir así tomar las decisiones sobre unos pocos grupos en vez de sobre cada uno de los individuos.

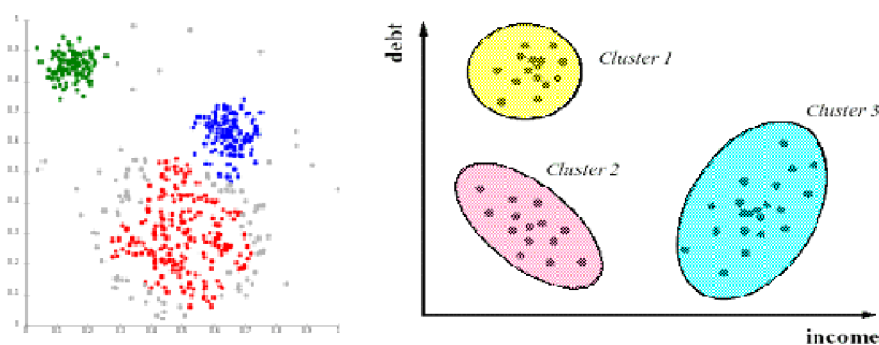
- El número de individuos puede ser muy elevado.
- Los atributos de cada individuo (“Me gusta”, seguidores, etc.) pueden ser muy heterogéneos.
- No se dispone de información sobre a qué grupo pertenece cada individuo.

Para facilitar la interpretación y la toma de decisiones, los algoritmos de agrupación acostumbran a representar los *clusters* de forma visual, a menudo diferenciados con colores, y acompañándolos de una leyenda que muestre aquellas características que definen a cada grupo (edad entre 20 y 30 años, más de 30 “Me gusta” en contenidos sobre libros de misterio, etc.).

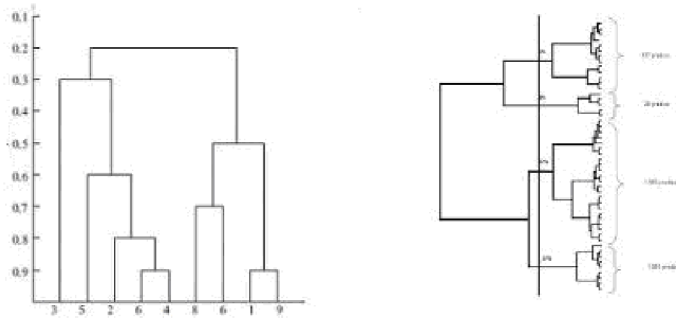
Hay dos tipos de representaciones comunes para las agrupaciones en redes sociales, disponibles en función del algoritmo de agrupación que se utilice:

1) **Agrupaciones o *clusters* planos.** Permiten diferenciar los grupos y los individuos que pertenecen a cada grupo, pero no se da ninguna información sobre la similitud o distancia que hay entre cada uno de los individuos que forman un grupo. La figura 9 muestra ejemplos de representación de *clusters* planos.

Figura 9. Representaciones de *clusters* planos



2) **Clustering jerárquico.** Se ordenan los individuos para que cada uno de ellos y cada nuevo grupo que se cree esté situado al lado de aquel que le es más similar. En la figura 10 podemos ver un par de ejemplos de *clusters* jerárquicos.

Figura 10. Representaciones de *clustering* jerárquicos

Los algoritmos de *clustering* pueden llegar a ser demasiado lentos si la cantidad de datos a utilizar, como sucede en redes sociales, es muy elevada. Por ello se deben utilizar algoritmos que finalicen con el mínimo número de iteraciones posibles sobre el conjunto de datos. Uno de los algoritmos de *clustering* más utilizados en las redes sociales es el de K-Means. Eso es debido a que, en redes sociales, suele ser más importante clasificar a un usuario en un grupo (usuario interesado en películas de suspense, románticas, de terror).

El algoritmo de K-means es un algoritmo de aprendizaje no supervisado que sirve para formar *clusters* (o grupos) de individuos similares. Es muy popular y utilizado debido a su simplicidad y rapidez de cálculo, hecho que le permite trabajar con grandes volúmenes de datos.

Para su funcionamiento se debe informar, a priori, de la cantidad de grupos que se quiere formar. El primer paso consiste en formar K grupos de un individuo escogido al azar, y clasificar el resto de usuarios en aquel grupo con el que tiene menor distancia. En el siguiente paso, se calcula el nuevo centro del grupo como la posición media de todos los usuarios que lo forman, y se clasifica de nuevo cada usuario en el grupo que tiene la posición media más cercana a él. Este paso se repite hasta que ningún usuario cambia de grupo respecto a la iteración anterior.

6.2. Clasificación de individuos

En algunas ocasiones se tendrá que decidir en qué grupo se clasifica un nuevo usuario o si se le recomienda un elemento o persona (sugerencias de amigos, recomendaciones de películas o música, si se le muestra un anuncio u otro, etc.).

En este tipo de casos, estamos ante un problema donde, además de los atributos que definen al usuario, tenemos la solución de casos anteriores en los que se ha planteado el mismo problema. Las soluciones de los casos anteriores han sido aportadas en muchos casos por los usuarios de la red social, ya sea otor-

gando una puntuación al elemento, o indicando si un elemento le gusta o no. En este tipo de problemas, al disponer de las soluciones de casos anteriores, se podrán utilizar técnicas de aprendizaje supervisado como:

- *Instance based-learning* (IBL: Aprendizaje basado en casos).
- Árboles de decisión.
- Redes neuronales.
- SVM (*Support Vector Machines*).
- Clasificadores bayesianos.
- Heurísticas y algoritmos genéticos.

Ved también

Para más información sobre estas técnicas se recomienda revisar los materiales de *Business Analytics*.

Al ser la cantidad de algoritmos a utilizar mucho más amplia que en los problemas de clasificación, se deberá evaluar el rendimiento de cada algoritmo, y escoger aquel que tenga una mayor capacidad de predicción. Para ello existen programas como Weka o R-Project que permiten ejecutar los algoritmos de clasificación supervisada más utilizados.

Tal y como sucedía en los algoritmos de *clustering*, en la elección de un algoritmo de clasificación para redes sociales, es de vital importancia que sea rápido, que sea fácil de integrar en la aplicación y que dé unos buenos resultados. Entre la gran variedad de algoritmos de clasificación disponibles, el algoritmo de los *k*-vecinos más próximos es uno de los más utilizados, ya que es uno de los que mejor cumple estos requisitos y es fácilmente integrable.

El algoritmo de los *k*-vecinos más próximos (*K-Nearest neighbour* en inglés) es un algoritmo de aprendizaje supervisado que clasifica un nuevo elemento en función de los otros elementos ya existentes en nuestro conjunto de datos. Para ello, se calcula la distancia del nuevo elemento con los otros elementos que ya forman parte del conjunto de datos. Aquel que tenga una distancia menor será considerado el vecino más próximo. El nuevo elemento se clasificará en el mismo grupo/resultado que el de su vecino más próximo.

A menudo, en vez de un solo elemento se utilizan los *k*-vecinos más próximos, siendo *k* normalmente un número impar. En dicho caso, el nuevo elemento se clasificará con el valor mayoritario entre sus vecinos más próximos.

6.2.1. Ejemplo de clasificador de la intención de usuario

En este ejemplo veremos cómo utilizar técnicas de clasificación para clasificar un texto de las redes sociales en tres categorías diferentes, que indican la polaridad de la opinión vertida en el texto: positivo, negativo y neutro.

Los textos son datos no estructurados que por sí solos no tienen ningún valor debido a su difícil interpretación y ambigüedad. Antes de poder tratar el texto, deben aplicarse técnicas de procesamiento de lenguaje natural que permitan “entender” su contenido en cierto sentido, y con un cierto margen de error. De estas técnicas, en redes sociales destacan las técnicas automáticas para detectar la intención de usuario³, (positiva, negativa o neutra) y aquellos que detectan el tipo de mensaje (opinión, pregunta, duda, queja, publicidad, mensaje personal, etc.).

⁽³⁾Esta intención puede entenderse como el sentimiento u opinión del usuario en el momento de escribir un texto. Más información sobre el tema en los siguientes artículos: <http://cs.stanford.edu/people/alecmgo/papers/TwitterDistantSupervision09.pdf>, <http://ravikiranj.net/drupal/201205/code/machine-learning/how-build-twitter-sentiment-analyzer>.

Estas técnicas son cada vez más populares en el tratamiento de información de redes sociales porque permiten seleccionar solo el conjunto de mensajes relevantes para el objetivo que estamos persiguiendo. Si, por ejemplo, queremos analizar los servicios o productos de nuestra empresa que tienen un mayor número de clientes insatisfechos, seleccionaremos solo aquellos de carácter negativo. En cambio, si queremos saber qué servicios o productos tienen más éxito en nuestros clientes para potenciarlos aún más, analizaremos los mensajes positivos.

A continuación utilizaremos un clasificador sencillo⁴ que, a partir del análisis de un conjunto inicial de mensajes, será capaz de detectar automáticamente una intención de usuario positiva, negativa o neutra en cada nuevo mensaje que se analice.

⁽⁴⁾Este sería el clasificador más sencillo sin ninguna optimización. En los artículos anteriores utilizan un esquema similar mejorando cada paso.

Este clasificador seguirá los siguientes pasos:

- En primer lugar se recuperará un número elevado de mensajes de Twitter y se clasificarán como positivos, negativos o neutros. Esta clasificación se deberá hacer manualmente, mensaje por mensaje. También está la opción de utilizar un conjunto de datos ya clasificados.

Webs recomendadas

Existen corpus de datos con mensajes clasificados para la detección de la intención de usuario. En inglés están disponibles en <http://www.sananalytics.com/lab/twitter-sentiment/> y <http://inclass.kaggle.com/c/si650winter11> y en español en <http://www.daedalus.es/TASS2013/corpus.php>.

- En segundo lugar se eliminarán las *stop words*⁵ de los mensajes. Una vez eliminadas se contará el número de veces que aparece cada palabra en mensajes positivos, negativos y neutros.
- A partir de este recuento se calculará la probabilidad de que la palabra pertenezca a un mensaje positivo, negativo o neutro. Si una palabra aparece 14 veces en mensajes positivos, 4 veces en mensajes negativos y 2 veces en mensajes neutros, entonces la probabilidad de que esta palabra denote un mensaje positivo será de un 0,7, de un 0,2 para un mensaje negativo y de un 0,1 para uno neutro.

⁽⁵⁾Una *stop word* es una palabra que no aporta significado en una cadena textual y que se elimina de la cadena de texto para que no influya en el resultado. Las preposiciones y las conjunciones son ejemplos de *stop words*.

Una vez calculadas las probabilidades de todas las palabras del conjunto de mensajes, el sistema ya será capaz de clasificar la intención de los nuevos mensajes. La tabla 1 muestra el ejemplo de calcular las probabilidades para las palabras *aprobaremos*, *todos* y *asignatura*.

Tabla 1. Recuento y las probabilidades de la intención de usuario para varias palabras

Palabra	Recuento en mensajes positivos	Recuento en mensajes negativos	Recuento en mensajes neutros	Probabilidad positivo	Probabilidad negativo	Probabilidad neutro
Aprobaremos	14	4	2	0.7	0.2	0.1
Todos	20	5	15	0.5	0.125	0.375
Asignatura	20	5	25	0.4	0.1	0.5

Una vez tenemos la tabla calculada, si llega un nuevo mensaje de texto podemos calcular la probabilidad de que este manifieste una opinión positiva, negativa o neutra. Para hacerlo, se calculará la probabilidad conjunta para cada una de las intenciones del mensaje. El mensaje será clasificado según la intención que obtenga una mayor probabilidad.

Por ejemplo, supongamos que nos llega el texto “aprobaremos todos la asignatura”. El cálculo de probabilidades conjuntas será el siguiente:

- $\text{Prob}(\text{aprobaremos todos la asignatura} \mid \text{positivo}) = 0.7 * 0.5 * 0.4 = 0.14$
- $\text{Prob}(\text{aprobaremos todos la asignatura} \mid \text{negativo}) = 0.2 * 0.125 * 0.1 = 0.0025$
- $\text{Prob}(\text{aprobaremos todos la asignatura} \mid \text{neutro}) = 0.1 * 0.375 * 0.5 = 0.01875$

Al ser la probabilidad de positivo mayor al resto de probabilidades, el mensaje sería clasificado como positivo. Notad que el sistema no es exacto, por lo tanto, en estos sistemas deberá calcularse y tenerse en cuenta el grado de fiabilidad del sistema, que dependerá en gran medida del método utilizado, y del número y calidad de los datos de entrenamiento (las clasificaciones de los mensajes realizadas).

6.3. Similitud entre elementos

Tanto los algoritmos de agrupación como los de clasificación utilizan a menudo funciones que calculan la distancia entre elementos. Para calcular la distancia entre dos elementos hay varias métricas que pueden aplicarse. Aplicar una u otra dependerá de varios factores, pero un factor importante es la naturaleza de los datos de los elementos (booleanos, enteros, decimales, geográficos, categóricos, etc.).

Probabilidad conjunta

La probabilidad conjunta se calcula multiplicando la probabilidad de cada una de las palabras. En caso de que una palabra no tenga una probabilidad asociada, se le dará un valor por defecto muy cercano a 0 para evitar que la probabilidad de todo el conjunto sea 0 debido a una sola palabra.

La cantidad de métricas diferentes para calcular la similitud entre elementos es muy grande. Debido a la gran cantidad de datos y a la necesidad de rapidez de cálculo, podemos destacar la distancia euclídea, el coeficiente de Sorensen-Dice, la distancia de Jaccard, la distancia de Hamming, o la distancia de Levenshtein. A continuación hacemos una pequeña introducción de estas métricas.

6.3.1. Distancia euclídea

La distancia euclídea se utiliza para calcular la distancia entre dos elementos que tienen atributos numéricos, y se define como la raíz cuadrada de la suma de las distancias al cuadrado de cada uno de los atributos que forman los elementos.

$$d_E(P_1, P_2) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (1)$$

A continuación mostramos cómo calcular la distancia euclídea mediante un ejemplo.

Supongamos que tenemos datos de la puntuación que han realizado tres usuarios sobre tres elementos:

Tabla 2. Puntuaciones otorgadas por usuarios a diferentes elementos

	Elemento 1	Elemento 2	Elemento 3
Usuario 1	0,1	0,5	0,9
Usuario 2	0,2	0,6	0,4
Usuario 3	0,3	0,5	0,7

A continuación podemos comprobar cómo se calculan las distancias entre los distintos usuarios, utilizando la diferencia de sus puntuaciones como distancia.

$$d_E(U_1, U_2) = \sqrt{(0.1 - 0.2)^2 + (0.5 - 0.6)^2 + (0.9 - 0.4)^2} = \sqrt{(-0.1)^2 + (-0.1)^2 + (0.5)^2} = \sqrt{0.045} = 0.2121 \quad (2)$$

$$d_E(U_1, U_3) = \sqrt{(0.1 - 0.3)^2 + (0.5 - 0.5)^2 + (0.9 - 0.7)^2} = \sqrt{(-0.2)^2 + (0)^2 + (0.2)^2} = \sqrt{0.08} = 0.2828 \quad (3)$$

$$d_E(U_2, U_3) = \sqrt{(0.2 - 0.3)^2 + (0.6 - 0.5)^2 + (0.4 - 0.7)^2} = \sqrt{(-0.1)^2 + (0.1)^2 + (0.3)^2} = \sqrt{0.11} = 0.3317 \quad (4)$$

6.3.2. Coeficiente de Sorensen-Dice y distancia de Jaccard

La distancia de Jaccard y el coeficiente de Sorensen-Dice son útiles en casos donde el número de atributos de cada individuo es diferente.

El coeficiente de Sorensen-Dice se define como:

$$QS = \frac{2C}{A+B} = \frac{2|A \cap B|}{|A| + |B|} \quad (5)$$

Donde $|A \cap B|$ son los atributos en común que tienen el elemento A y el elemento B.

- $|A|$ es el número de atributos de A
- $|B|$ es el número de atributos de B

A partir del coeficiente de Sorensen-Dice, se define la distancia de Jaccard entre elementos como:

$$d = 1 - \frac{2|X \cap Y|}{|X| + |Y|} \quad (6)$$

Pongamos, por ejemplo, una tienda *online* con una alta rotación de productos (películas, libros, dispositivos de electrónica etc.) y que publica en Facebook sus productos, sobre los cuales los usuarios pueden hacer “Me gusta”. Supongamos que queremos crear grupos de usuarios para enviarles ofertas personalizadas sobre una misma temática o sobre temáticas relacionadas a partir de los “Me gusta” que hayan realizado. Si utilizáramos el coeficiente de Sorensen-Dice para calcular la distancia entre elementos, podríamos definir un atributo como un “Me gusta” realizado por un usuario U sobre un producto P . Por lo tanto, dados dos usuarios U_a y U_b , $|A|$ sería el número de “Me gusta” del usuario U_a , $|B|$ sería el número de “Me gusta” del usuario U_b y $|A \cap B|$ sería el número de “Me gusta” de ambos usuarios sobre el mismo producto.

Utilizar el coeficiente de Sorensen-Dice o la distancia de Jaccard en este caso podría ser conveniente ya que:

- Los usuarios probablemente tendrán diferentes números de atributos, es decir, han realizado un número de “Me gusta” diferente.
- La distancia entre dos usuarios será pequeña cuantos más elementos (por ejemplo, “Me gusta”) tengan en común.
- Si un usuario hace pocos “Me gusta” y un segundo usuario hace muchos, la distancia entre ellos será mayor a si los dos hacen un número parecido de “Me gusta”. Esto es debido a que esta medida de distancia tiene en cuenta el número de atributos de cada elemento, en este caso el número de “Me gusta”.

6.3.3. Distancia de Hamming

La distancia de Hamming se utiliza para calcular la distancia entre elementos formados por variables booleanas o categóricas de igual longitud o con el mismo número de atributos. La distancia entre ellos consiste en el número de atributos diferentes entre los dos elementos. También es una distancia apropiada para calcular distancias entre textos de igual longitud.

A modo de ejemplo, la distancia de Hamming entre las palabras *casal* y *pasar* es de 2.

CASAL

PASAR

6.3.4. Distancia de Levenshtein

La distancia de Levenshtein entre dos secuencias de elementos se calcula como el número mínimo de operaciones requeridas para transformar una secuencia en la otra.

Por ejemplo, la distancia entre las palabras *capa* y *cala* será 1, ya que solo es necesario sustituir la 'p' por la 'l'. Por otro lado, la distancia entre *casa* y *calle* será 3 porque se necesitan al menos tres cambios para convertir un elemento en otro: 1) sustitución de 's' por 'l', 2) inserción de la segunda 'l' y 3) sustitución de 'a' por 'e'.

Esta distancia es muy utilizada en redes sociales debido a que permite calcular la distancia entre usuarios o elementos que contienen un número de atributos diferentes.

La distancia de Levenshtein es utilizada a menudo sobre la transcripción fonética de una palabra con el algoritmo Metaphone. Este algoritmo aplica un sencillo conjunto de reglas de sustitución para realizar su transcripción fonética, y minimiza la distancia de dos palabras que, pese a escribirse de diferente manera, se pronuncian de forma similar. Esta transformación puede ser utilizada para calcular distancias entre textos, minimizando el efecto de palabras escritas con faltas ortográficas.

6.4. Teoría de grafos

Un grafo es un conjunto de vértices V (también llamados nodos) y un conjunto de aristas entre ellos. Si existe una arista $\{a, b\}$ entre los nodos a y b , entonces podemos decir que los nodos a y b están relacionados.

Las aristas pueden ser:

- **Aristas sin dirección.** Una arista $\{a, b\}$ relaciona los nodos a y b en ambas direcciones.
- **Aristas dirigidas.** Una arista $\{a, b\}$ relaciona los nodos en la dirección de a hacia b , pero no en la dirección de b hacia a .

Los grafos dirigidos son grafos que solo contienen aristas dirigidas.

Hay muchas situaciones cotidianas que pueden representarse mediante grafos, como por ejemplo la web. En este caso, las páginas web serán los vértices, y los enlaces (*links*) determinarán las aristas. Como los enlaces web son unidireccionales, el grafo resultante será dirigido. Esta es la base del algoritmo de *page rank* de Google, donde la importancia de un nodo o página web viene determinada por el número de sitios que enlazan a esta web, ponderado por la importancia de la página web de la que sale el enlace.

Como se puede ver en la figura 10, la estructura de las redes sociales también se puede representar mediante grafos.

Las redes sociales que conocemos hoy en día no pueden representarse como una red global, sino como un conjunto de subredes conectadas entre sí mediante unos nodos de intermediación, permitiendo que desde cualquier elemento de una subred se pueda llegar a cualquier otra subred con pocas conexiones.

Los subgrupos de una red social tienden a estar fuertemente conectados entre sí. Es decir, dos nodos diferentes en un mismo subgrupo acostumbran a tener muchas relaciones en común con los otros nodos del mismo subgrupo (mis amigos y yo acostumbramos a tener muchos amigos en común). La creación de estos subgrupos y los puentes que existen entre ellos son los que permiten aplicar técnicas para encontrar grupos con características similares (segmentar el mercado) y descubrir aquellos nodos (o usuarios) capaces de llegar a muchos grupos diferentes (líderes de opinión o embajadores).

La teoría de grafos no se aplica solo a usuarios, sino que se puede utilizar en muchos ámbitos. Por ejemplo, dado el título de un libro sería posible utilizar grafos para descubrir y representar los 100 recursos y/o agentes más relevantes relacionados con dicho libro.

Vídeo

En el vídeo <http://vimeo.com/32559678> se puede ver cómo, a partir del título de un libro disponible en Amazon y de la herramienta de visualización de grafos *gephi* (<http://gephi.org>), se crea la red de los 100 elementos más importantes que lo rodean.

7. Sistemas de recomendación

Podemos definir un sistema de recomendación como un sistema inteligente que proporciona a los usuarios una serie de sugerencias personalizadas (llamadas recomendaciones) sobre un determinado tipo de elementos. Los sistemas de recomendación estudian las características y los gustos de cada usuario y, mediante un procesamiento de los datos o mediante el cálculo de distancias, encuentran un subconjunto de elementos que pueden resultar de interés para el usuario. En el caso de las redes sociales, los elementos a recomendar pueden ser variados e ir desde productos o servicios hasta usuarios, pasando por localizaciones geográficas.

Los sistemas de recomendación en redes sociales verticales están más maduros que los de las redes sociales horizontales, dado que, al centrarse la actividad de la red social sobre un solo tema de interés (noticias, vídeos y series *online*, libros, etc.), su contexto es más acotado y fácil de tratar. En las redes sociales horizontales, para cada posible contexto en el que se puedan realizar recomendaciones, se debe hacer una selección de elementos y de individuos a los que se les harán las recomendaciones antes de aplicar este sistema.

Algunas redes sociales verticales han conseguido la mayor parte de éxito gracias a sistemas de recomendación. Por ejemplo, las redes sociales que comparten artículos de diarios o *blogs*, como Reddit o Meneame, seleccionan las noticias que aparecen en su portada o en la página principal de cada categoría a partir de las votaciones realizadas por los usuarios. En cambio, otras redes sociales sobre visualización de contenido multimedia, como pueden ser YouTube o Flixster, basan las recomendaciones realizadas a un usuario en función de aquellos usuarios que tienen un gusto más parecido, y no a partir de las votaciones de todos los usuarios.

En redes sociales generales como Facebook o Twitter, pese a ser más complejo realizar recomendaciones, también se utilizan los sistemas de recomendación para proponer amigos o personas a las que seguir. Las redes sociales generales, y sobre todo Facebook, tienen un gran potencial como sistemas de recomendación. A partir de los “Me gusta” indicados por los usuarios, por ejemplo, se puede proponer conocer a personas afines para acudir a determinados eventos o espectáculos.

Los sistemas de recomendación se pueden clasificar en:

- **Sistemas de recomendación basados en el contenido:** Las recomendaciones se generan a partir de características de los elementos a recomendar y a la información aportada por otros usuarios sobre su contenido. Con esta información, el sistema aprende a clasificar los elementos en función

de las valoraciones de otros usuarios. Un ejemplo de sistema de recomendación basado en contenido es el recomendador de Youtube. Este utiliza etiquetas y metadatos de los vídeos visionados por un usuario por un lado, y las etiquetas y los metadatos de los vídeos potencialmente recomendables por el otro. Comparando esta información, Youtube recomienda al usuario un conjunto de vídeos basados en su historial de visualizaciones.

- **Sistemas de recomendación demográficos:** Las recomendaciones se basan en perfiles demográficos de usuarios (edad, sexo, profesión, etc.). El sistema de recomendación crea una clasificación inicial basada en el perfil demográfico del usuario y después la mejora con sus votaciones. Existe un buen número de aplicaciones móviles que recomiendan a usuarios en función de la edad, el sexo, su geolocalización y su disponibilidad para ir a tomar una copa o acudir a algún evento.
- **Sistemas de recomendación basados en el conocimiento:** Las recomendaciones se generan según las inferencias que el sistema realiza acerca de las necesidades y preferencias de los usuarios. Estas inferencias son generalmente realizadas con alguna lógica adicional, con el fin de que coincidan las características del elemento con las necesidades del usuario. Amazon utiliza este tipo de sistema de recomendación. Cuando un usuario se da de alta en su web, el sistema registra sus páginas vistas, los productos que ha añadido al carrito de la compra y las compras que realiza. A partir de esta información, un algoritmo es capaz de inferir qué productos pueden ser de interés para el usuario. Su sistema de recomendación de libros se basa en el algoritmo de recomendación creado por **Goodreads**, que recomienda libros a los usuarios a partir de las siguientes cuatro características:
 - **Catalogación:** Sistema de puntuación que permite al usuario otorgar a cada libro una calificación de entre una y cinco estrellas, con la opción de acompañarla con una opinión escrita. Esta opinión escrita puede ser analizada para extraer y tener en cuenta los sentimientos u opiniones que contiene.
 - **Comunidad de usuarios:** Lectores, grupos de lectores con gustos similares y los autores de los libros.
 - **Recomendaciones:** Basa sus recomendaciones a partir de las puntuaciones otorgadas a otros libros leídos por los usuarios.
 - **Títulos:** Se puede encontrar prácticamente cualquier libro que se haya editado. Los que no se encuentran pueden ser añadidos por los usuarios.
- **Sistemas de recomendación colaborativos:** En estos sistemas las recomendaciones son generadas usando exclusivamente información proporcionada por los usuarios. En los sistemas de recomendación colaborativos se buscan usuarios con un historial de votaciones similares, y se generan

Goodreads

Goodreads es un portal de recomendación de libros que salió a la luz en el 2006 y en el 2013, tenía 11 millones de usuarios, 395.000 libros puntuados y más de 20.000 clubs de lectura creados por sus propios usuarios. <http://www.goodreads.com/>.

las recomendaciones para un usuario a partir de los elementos votados por sus vecinos, pero que no han sido todavía votados por el usuario.

- Destacan dos tipos de sistema de recomendación colaborativos:
 - Los **métodos basados en el conocimiento** utilizan una matriz de votaciones creada a partir de los elementos votados por los usuarios y utilizan algoritmos de inteligencia artificial, como redes neuronales, clasificadores bayesianos, sistemas difusos o algoritmos genéticos para calcular la similitud entre usuarios.
 - Los **métodos basados en memoria** utilizan métricas de similitud entre usuarios para determinar el parecido de un par de usuarios entre sí. La similitud se calcula a partir de los elementos votados por ambos usuarios y la nota otorgada por cada uno de ellos a cada elemento.

En la actualidad es común encontrar sistemas de recomendación híbridos que combinan varias de las técnicas mencionadas.

7.1. Ejemplo de sistemas de recomendación de películas

En el siguiente ejemplo veremos cómo realizar un sistema de recomendación colaborativo basado en memoria utilizando el algoritmo de los k -vecinos. Imaginemos una página web de visionado de películas y series en *streaming*. El sitio web tendrá miles de vídeos y series (los elementos) en su base datos y una base de datos de miles de usuarios registrados a los que deberá hacer recomendaciones. Los usuarios votarán, con valores del 1 al 10, las películas o series que han visto. A partir de estas votaciones, se calcularán las similitudes entre usuarios y se recomendarán nuevos elementos.

Con las votaciones de los usuarios creamos la matriz de votaciones, donde:

- U_x representa a un usuario.
- E_y representa a una película o serie.
- Un valor numérico en una celda de la matriz indica la puntuación que el usuario U_x ha otorgado al elemento E_y .
- Una casilla vacía en la matriz indica que el usuario no ha valorado el elemento.

Tabla 3. Votaciones de películas realizadas por usuarios

	E₁	E₂	E₃	E₄	E₅	E₆	E₇	E₈	E₉	E₁₀	E₁₁	E₁₂
U₁	2			4	4		1		2	3		1
U₂		5	5	1	5	1		1		3		2

	E ₁	E ₂	E ₃	E ₄	E ₅	E ₆	E ₇	E ₈	E ₉	E ₁₀	E ₁₁	E ₁₂
U ₃		2	1	3			1				4	
U ₄	1	4			4	1		4		3	5	
U ₅						1	2	5	1	1		2
U ₆		2		4	1	2		2	1			

Una vez creada la matriz, se calculará la distancia entre cada pareja de usuarios utilizando la diferencia cuadrática media, teniendo en cuenta únicamente aquellos elementos valorados por ambos usuarios:

$$im(x, y) = 1 - \frac{1}{\# B_{x,y}} \sqrt{\sum_{i \in \text{Elementos}} \left(\frac{r_{x,i} - r_{y,i}}{\max - \min} \right)^2} \in [0, 1]$$

#B_{x,y} es el número de elementos que ambos usuarios han votado.

max es la puntuación máxima que un usuario ha otorgado a un elemento.

min es la puntuación mínima que un usuario ha otorgado a un elemento.

r_{x,i} es la puntuación que el usuario U_x ha otorgado al elemento E_i.

r_{y,i} es la puntuación que el usuario U_y ha otorgado al elemento E_i.

Entre los usuarios U₁ y U₂ la distancia es:

$$sim(U_1, U_2) = 1 - \frac{1}{4} \cdot \sqrt{\left(\frac{4-1}{5-1}\right)^2 + \left(\frac{4-5}{5-1}\right)^2 + \left(\frac{3-3}{5-1}\right)^2 + \left(\frac{1-2}{5-1}\right)^2} = 0,739 \quad (8)$$

Una vez calculados todos los pares de distancias entre usuarios, obtendremos la siguiente matriz de similitud de usuarios:

Tabla 4. Distancia entre usuarios

Sim(x, y)	U ₁	U ₂	U ₃	U ₄	U ₅	U ₆
U ₁		0.793	0.875	0.917	0.835	0.737
U ₂			0.551	0.834	0.721	0.7
U ₃				0.721	0.485	0.875
U ₄					0.814	0.735
U ₅						0.737
U ₆						

Diferencia cuadrática media

La diferencia cuadrática media es una generalización de la distancia euclídea. En lugar de devolver la distancia entre dos elementos, devuelve la similitud entre ellos. La similitud se define como:
 Similitud = 1 - Distancia

El tercer paso consiste en ordenar las similitudes entre usuarios de mayor a menor, y escoger los k -vecinos más próximos a cada usuario. Con $k=2$ tendremos:

Tabla 5. Conjunto de los usuarios más similares a cada uno de los usuarios

$k=2$	U_1	U_2	U_3	U_4	U_5	U_6
K_u	{ U_4, U_3 }	{ U_4, U_1 }	{ U_6, U_1 }	{ U_1, U_2 }	{ U_1, U_4 }	{ U_3, U_5 }

El cuarto paso consiste en estimar la puntuación que el usuario U_x otorgaría a cada elemento E_x en función de la puntuación otorgada por sus k -vecinos más próximos. La puntuación de un elemento E_i para un usuario U_x se puede estimar a partir de la media aritmética de las puntuaciones otorgadas por sus k -vecinos más próximos.

La puntuación estimada del usuario U_1 para el elemento E_2 en función de sus 2 vecinos más próximos (U_4, U_3), utilizando la media aritmética es:

$$U_1, E_2 = \frac{1}{2} \cdot (U_4, E_2 + U_3, E_2) = \frac{1}{2} \cdot (4 + 2) = 3 \quad (9)$$

Y la matriz de puntuaciones estimadas a partir de los vecinos más próximos para cada usuario y elemento es:

Tabla 6. Predicción de la puntuación que un usuario otorgaría a una película

	E_1	E_2	E_3	E_4	E_5	E_6	E_7	E_8	E_9	E_{10}	E_{11}	E_{12}
U_1		3	1			1		4			4,5	
U_2	1,5						1		2		5	
U_3	2				2,5	2		2	1,5	3		1
U_4			5	2,5			1		2			1,5
U_5	1,5	4		4	4						5	
U_6			1				1,5			1	4	2

Las puntuaciones de un usuario U_x también se podrían estimar ponderando la votación del elemento E_i de cada usuario próximo, por el grado de similitud de este vecino con el usuario U_x .

Una vez que se han estimado las votaciones de los elementos que el usuario todavía no ha valorado, queda identificar los elementos potencialmente relevantes para cada usuario. Para ello se ordenarán de mayor a menor aquellos elementos para los que se ha podido calcular una predicción. El resultado lo podemos ver en la tabla 7. Por cada usuario, encontramos una lista de los elementos que aún no ha consumido, ordenados por su potencial interés. Por lo

tanto, para recomendar los dos elementos más relevantes para cada usuario, solo se deberá seleccionar los dos primeros elementos de cada secuencia, como se ve en la tabla 7 (Zu).

Tabla 7. Recomendaciones de elementos a cada usuario

N=2	U₁	U₂	U₃	U₄	U₅	U₆
Xu	{E ₂ ,E ₃ ,E ₆ ,E ₈ ,E ₁₁ }	{E ₁ ,E ₇ ,E ₉ ,E ₁₁ }	{E ₁ ,E ₅ ,E ₆ ,E ₈ ,E ₉ ,E ₁₀ ,E ₁₂ }	{E ₃ ,E ₄ ,E ₇ ,E ₉ ,E ₁₂ }	{E ₁ ,E ₂ ,E ₄ ,E ₅ ,E ₁₁ }	{E ₃ ,E ₇ ,E ₁₀ ,E ₁₁ ,E ₁₂ }
Zu	{E ₁₁ ,E ₈ }	{E ₁₁ ,E ₉ }	{E ₁₀ ,E ₅ }	{E ₃ ,E ₄ }	{E ₁₁ ,E ₄ }	{E ₁₁ ,E ₁₂ }

Resumen

En este módulo hemos visto que las redes sociales están presentes en gran parte de las tareas que realizamos a diario y que esta nueva forma de comunicación ha supuesto un aumento de la información generada, una inmediatez de comunicación y una mayor variedad en la forma y origen de las comunicaciones.

Gracias a la información obtenida en las redes sociales, es posible conocer mucho más en detalle las preferencias de los clientes, realizar algunas tareas de forma más eficiente y respaldar la toma de decisiones con información más precisa.

También hemos visto que el gran volumen de datos generado ha traído consigo la necesidad de utilizar nuevos sistemas de almacenamiento y nuevas herramientas de medición de información y generación de informes para facilitar la toma de decisiones.

Para aquellos casos donde las herramientas de medición no nos puedan facilitar la medición y seguimiento de los indicadores a evaluar, hemos introducido las tecnologías que nos permiten recuperar la información de las redes sociales, y algunas técnicas matemáticas y de minería de datos para elaborar soluciones personalizadas.

Por último, se han mostrado algunos ejemplos sobre cómo personalizar la información que se muestra a un usuario a partir de la información recogida de las redes sociales.

Bibliografía

[IEEE Engineering Management Review Volume 41, 3, Sept 2013]

<http://www.deepdyve.com/browse/journals/engineering-management-review-ieee>

[Alec Go; Richa Bhayani; Lei Huang. Twitter Sentiment Classification using Distant Supervision]

<http://cs.stanford.edu/people/alecmgo/papers/TwitterDistantSupervision09.pdf>

[www.thinknook.com Twitter Sentiment Analysis Training Corpus (Dataset)]

<http://thinknook.com/twitter-sentiment-analysis-training-corpus-dataset-2012-09-22/>

[D. S. Rajput; R. S. Thakur; G. S. Thakur & 4 NEERAJ Sahu. Analysis of Social Networking Sites Using K- Mean Clustering Algorithm]

http://www.idc-online.com/technical_references/pdfs/information_technology/Analysis%20of%20Social.pdf

[Manh Cuong Pham; Yiwei Cao; Ralf Klamma; Matthias Jarke. A Clustering Approach for Collaborative Filtering Recommendation Using Social Network Analysis]

<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.232.2716&rep=rep1&type=pdf>

[David Caldevilla Domínguez. Las Redes Sociales. Tipología, uso y consumo de las redes 2.0 en la sociedad digital actual]

<http://revistas.ucm.es/index.php/DCIN/article/viewFile/DCIN1010110045A/18656>

[Nina Mishra, Robert Schreiber; Isabelle Stanton; Robert E. Tarjan. Clustering Social Networks]

<http://www.cs.berkeley.edu/~isabelle/papers/cci.pdf>

[William B. Frakes; Ricardo Baeza-Yates. Information Retrieval: Data Structures & Algorithms,]

<http://orion.lcg.ufrj.br/Dr.Dobbs/books/book5/toc.htm>

[David Caldevilla Domínguez. Las Redes Sociales. Tipología, uso y consumo de las redes 2.0 en la sociedad digital actual]

<http://revistas.ucm.es/index.php/DCIN/article/viewFile/DCIN1010110045A/18656>

[DatKnoSys.com, BigData para analizar social media]

<http://www.slideshare.net/DatKnoSys/big-data-para-analizar-las-redes-sociales>

