

Enunciado

Àlex Bartrolí Muñoz

PID_00209808



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. Fundació 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. Análisis del contexto	7
1.1. Los <i>stakeholders</i> : actores vinculados a la actividad asistencial del hospital	7
1.2. El funcionamiento básico de las áreas de urgencias	9
1.3. El funcionamiento básico de las áreas de hospitalización	11
1.4. Enfermedades infecciosas	13
1.5. Redes sociales	15
2. Motivación	18
3. Enunciado del caso práctico	19
3.1. Trabajo que realizar	19
3.2. Fuentes de datos suministradas	21
3.3. Notas técnicas: Recuperación de datos de Twitter	21

Introducción

El contexto del material que presentamos a continuación está basado en un caso real, la evolución del modelo sanitario catalán y la transformación de sus sistemas de información para adaptarse al entorno y a las nuevas estrategias planteadas, descritas y desarrolladas en el material docente “Los sistemas de información en el Institut Català de la Salut. Trazando una estrategia BI a partir del Proyecto Argos”. Este material, titulado “Sistema de predicción de brotes virales a partir de la información de redes sociales”, es la continuación de los casos prácticos anteriores “Implementando un cuadro de mando integral en un hospital general básico” y “Diseño de un almacén de datos para la gestión de hospitalización de un hospital básico general” y está enfocado a la definición de un sistema de predicción de brotes virales de gripe a partir de la información introducida por los usuarios en las redes sociales.

El objetivo de este trabajo es desarrollar un sistema capaz de predecir la aparición de brotes víricos gripales antes de que se produzcan. Para ello se definirá un sistema capaz de recuperar de las redes sociales los mensajes publicados por los usuarios sobre la gripe y comparar el grado de similitud de la situación actual con periodos de años anteriores que han precedido a un brote gripal.

Con el fin de poder desarrollar un proyecto lo más específico posible, el estudiante deberá definir los procesos de recuperación de información de las redes sociales, elaborar un cuadro de mando para la toma de decisiones, definir una estrategia de almacenamiento y explotación de información del almacén de datos e indicar posibles mejoras futuras de la herramienta de predicción.

Cabe considerar que en circunstancias normales, incluso en los centros líderes en la adopción e implantación de herramientas de minería y explotación de la información, el desarrollo de estas herramientas supone proyectos de meses o incluso años, con la participación de equipos multidisciplinares que los van implementando a lo largo del tiempo, en un proceso de mejora continua.

En este sentido, el objetivo del caso no es tanto desarrollar exhaustivamente el sistema de predicción, sino entender el proceso de extracción de información de las redes sociales, sus limitaciones y aprender cómo analizar estos datos de forma efectiva, creando algunos artefactos de BI predictivos que puedan ser integrados en los cuadros de mandos y almacenes de datos trabajados anteriormente.

1. Análisis del contexto

El desarrollo de este caso se centra fundamentalmente en las áreas de urgencias y hospitalización. La previsión de brotes víricos puede permitir realizar una estimación de los servicios de urgencias y hospitalización y por tanto prevenir posibles aglomeraciones por retrasos en los tratamientos.

Para que el estudiante tenga una referencia más concreta sobre la que desarrollar el caso, se propone una estructura genérica de la organización, así como una descripción general de los procesos básicos de urgencias y hospitalización. Esta estructura se ofrece solamente a título orientativo sin que tenga que condicionar el desarrollo del caso. El propósito principal es que sirva como referencia a todos aquellos que no estén familiarizados con estos entornos, con el fin de que puedan comprender mejor la actividad propia de un centro hospitalario.

1.1. Los *stakeholders*: actores vinculados a la actividad asistencial del hospital

Antes de entrar a detallar el funcionamiento de las áreas de urgencias y hospitalización, haremos un breve análisis de los principales actores que se deben tener en cuenta en la definición de un sistema de detección de brotes víricos y, en consecuencia, qué tipo de preguntas debe ser capaz de responder nuestro sistema de predicción.

Los actores involucrados en un sistema de salud suelen tener intereses muy diversos y a menudo contrapuestos. En nuestro caso particular los actores que tener en cuenta son:

1) **Las autoridades sanitarias**, tanto por lo que respecta al Departamento de Salud de la Generalitat como al Catalunya. A partir de sus políticas sanitarias y de compra de servicios, determinarán de forma clave la actividad que debe proporcionar finalmente nuestro centro. No podemos olvidar que un altísimo porcentaje de la facturación de los hospitales públicos proviene de las propias autoridades sanitarias y, por lo tanto, será esencial dotarnos de mecanismos que nos aseguren que nuestra cartera de servicios y nuestra “capacidad de producción” están ajustadas a la oferta que debemos proveer.

2) **La institución ICS**. En nuestro caso, al tratarse de un centro del Instituto Catalán de la Salud (ICS), los objetivos estratégicos de la institución ICS necesariamente se trasladarán a cada uno de los centros. Con toda probabilidad, gran parte de estos objetivos estarán completamente alineados con los objetivos de las autoridades sanitarias, pero habrá otros que estarán condicionados

por las propias directrices corporativas o, incluso, por los procesos internos de gestión en ámbitos como pueden ser: logística, compra agregada, política de recursos humanos, compra de productos intermedios, etc.

3) La propia dirección del centro. Obviamente, tendrá un papel clave según el grado de autonomía que le otorga el propio ICS, supuestamente cada vez mayor debido a los procesos de descentralización que se están produciendo dentro del nuevo modelo de empresa pública, tal como ya apuntábamos en el desarrollo del primer caso. Probablemente, en un hospital general básico el grado de diferenciación o especialización no será significativo y, por tanto, el grado de personalización y especificidad de su sistema de cuadro de mando tampoco lo será.

4) Los profesionales. Otro de los actores relevantes serán los propios profesionales que desarrollan su actividad en las diferentes áreas, como asistencial, administrativa, etc. Será esencial para los responsables de las distintas áreas disponer de mecanismos para el seguimiento y control de sus actividades, así como de la calidad de los servicios que prestan, tanto desde un punto de vista cualitativo como cuantitativo. Hay que tener en cuenta que su actividad estará enormemente condicionada por la oferta que deben generar (de acuerdo con la compra de servicios pactada con las autoridades sanitarias) y de los recursos de que dispongan para satisfacerla de acuerdo con determinados estándares de calidad, propios o inducidos por otros actores.

5) El propio entorno sanitario del centro. Es imprescindible tener en cuenta la interrelación y colaboración entre los diferentes actores dentro de un mismo territorio (área capitativa).

6) El entorno político y social del centro. No debemos olvidar que estamos trabajando con servicios públicos y que se establece una estrecha relación entre los distintos ámbitos sociopolíticos locales: gobiernos locales, servicios sociales, agrupaciones de vecinos, colectivos específicos, etc. Aunque no existe una relación directa entre la atención individualizada y el papel que desempeñan estos colectivos, está claro que tienen un papel decisivo en lo referente a representación social y, por lo tanto, es necesario tenerlos en cuenta en el desarrollo de determinadas estrategias.

7) El ciudadano. El último actor que describir, aunque probablemente uno de los más relevantes. El ciudadano, entendido como el paciente y su entorno familiar directo, es el beneficiario directo del servicio prestado. Sobre él será preciso desarrollar un seguimiento específico, y no solo desde el punto de vista asistencial sino también desde el emocional y espiritual. En este sentido, cabe tener presente que la atención sanitaria va mucho más allá del acto meramente clínico y que comporta muchas otras variables, que a menudo pueden tener tanto o más peso que la propia curación (siempre que sea posible) en

la percepción del paciente: atención prestada, buen trato, respeto, celeridad, diligencia, cantidad y calidad de la información recibida, soporte emocional, psicológico, etc.

A menudo veremos que muchos de los objetivos estratégicos pueden estar relacionados con diferentes actores. Esto lo podemos observar rápidamente en los documentos de referencia de las autoridades sanitarias:

- “Pla de Salut 2011-2015”, del Departamento de Salud.
- “Tercer Informe” de la Central de Resultados.

En ellos podemos ver el grado de detalle y la diversidad de ámbitos sobre los que se recoge información, y que, cada vez de forma más clara, se reflejan en los contratos de servicios que se establecen entre las autoridades sanitarias y los distintos centros sanitarios.

1.2. El funcionamiento básico de las áreas de urgencias

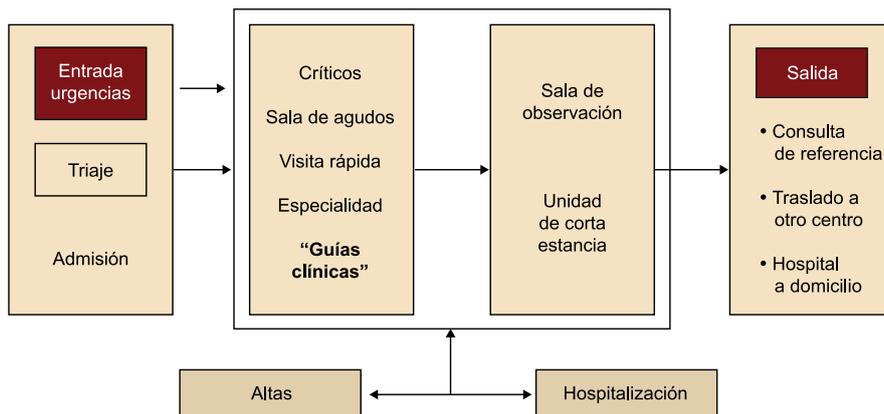
Las áreas de urgencias hospitalarias son las responsables de dar respuesta a las necesidades de salud de los ciudadanos, en lo que se refiere a la atención de urgencias y emergencias que tengan que ver con problemas de salud sobrevenidos, sea cual sea su causa.

Normalmente, estas áreas están diseñadas para la atención de urgencias críticas y graves. Aunque históricamente, debido a diferentes motivos, como la facilidad de acceso o el desconocimiento, han sido utilizadas por parte de la mayoría de los ciudadanos para el tratamiento de cualquier problema de salud imprevisto o sobrevenido, en lugar de dirigirse a la atención primaria. Este hecho ha impactado enormemente en su funcionamiento y ha generado efectos negativos, como la sobreocupación de espacios, las largas esperas, el hacinamiento, los tiempos de asistencias anormalmente largos, los tiempos de espera excesivos para la realización de pruebas complementarias, etc. Todo ello ha implicado una serie de problemas de gestión y ha provocado una pérdida de calidad del funcionamiento de las áreas de urgencia.

Además, tal y como apuntábamos en el caso práctico anterior, la enorme variabilidad de la práctica clínica en el uso de los recursos para cada una de las asistencias puede ser tremendamente cambiante.

A continuación mostramos un circuito estándar de atención en urgencias.

Figura 1. Funcionamiento general de un área de urgencias



Los principales canales de entrada a las áreas de urgencias suelen ser los siguientes:

- El desplazamiento del propio paciente al servicio de urgencias por sus propios medios.
- A través de transporte sanitario, tras un accidente laboral o de tráfico o de otras circunstancias, que le impiden la movilidad.
- La derivación desde otros centros sanitarios o de atención primaria, por saturación o por no disponer de las especialidades y/o medios necesarios, no pueden atender al paciente en origen.

Dependiendo de los síntomas, el diagnóstico inicial y la severidad potencial de la urgencia detectada en las unidades de triaje, el paciente será asignado a una unidad para su tratamiento y seguimiento, con el fin de determinar su diagnóstico y dar solución al problema de salud que presenta.

Unidades de triaje

En las unidades de triaje se establece una primera clasificación de los pacientes en función de su criticidad y patología. La atención de los pacientes se priorizará por las decisiones de triaje y no por orden de llegada. El objetivo es atender rápidamente a los pacientes críticos.

Un servicio de urgencias, como el de nuestro caso, generalmente suele contar con una serie de servicios o especialidades básicas, ya que dispone de recursos humanos y técnicos especializados para el tratamiento de problemas de salud relativos a los mismos. A modo de ejemplo, los servicios básicos pueden ser:

- Urgencias generales.
- Urgencias obstétricas (relativas al parto) y ginecológicas.
- Urgencias pediátricas.
- Urgencias traumatológicas.
- Urgencias oftalmológicas.

Adicionalmente podremos encontrar:

- Las unidades o salas de observación, que no están vinculadas necesariamente a una especialidad sino que suelen ser espacios polivalentes, en los cuales el paciente queda en observación con el fin de evitar la ocupación de las áreas específicas de cada servicio, o asegurar que pueden ser fácilmente supervisados por personal médico.
- Las unidades de corta estancia, que están dedicadas a los tratamientos que no requieren de hospitalización, pero que dada su urgencia no suelen tratarse dentro de las áreas de consulta externa.

Normalmente, el paciente será sometido a pruebas diagnósticas (radiología, laboratorio, diagnóstico por imagen, etc.), a tratamientos farmacológicos y a los procedimientos médicos y quirúrgicos que se consideren necesarios en función de la patología diagnosticada.

Las asistencias en urgencias derivan en uno de los siguientes motivos de alta:

- Alta a domicilio (normalmente, según criterio médico, aunque también puede ser por abandono o voluntaria).
- Ingresos (con o sin intervención).
- Derivación hacia otro centro sanitario.
- Defunción.

Además de los requerimientos que se puedan establecer desde el propio centro o urgencias, el Calsalut también monitoriza, especialmente en determinadas épocas del año o zonas, el funcionamiento de dichas áreas con el fin de identificar situaciones de excesiva saturación o incluso analizar posibles problemas epidemiológicos que puedan surgir y que afecten a la salud pública. Es en este punto donde el uso de técnicas de BI social pueden ser útiles para estimar potenciales brotes virales y, así, ser capaces de dimensionar los servicios de urgencias en función de estas previsiones.

En este sentido, a través del PIUC (Plan integral de urgencias de Cataluña) se monitorizan a diario determinados parámetros de funcionamiento y clínicos de las áreas de urgencias de todos los centros de la XHUP2 (Red hospitalaria de utilización pública).

1.3. El funcionamiento básico de las áreas de hospitalización

Los hospitales están organizados (desde un punto de vista asistencial) en servicios clínicos, con mayor o menor nivel de especialización dependiendo de su categoría.

De forma genérica, los servicios clínicos se organizan en dos grandes ámbitos, el médico y el quirúrgico, que se diferencian fundamentalmente en el uso que se realiza de las áreas quirúrgicas (preanestesia, quirófanos y reanimación) por parte de los médicos.

Aunque esta separación no siempre es completamente clara, ya que hay subespecialidades dentro de la cartera de servicios que superponen los dos ámbitos, los servicios habituales en un centro hospitalario básico son los siguientes:

Tabla 1. Esquema general de organización de servicios hospitalarios (en hospitalización)

Servicios médicos	Servicios quirúrgicos
Medicina interna	Cirugía general y digestiva
Cardiología	Cirugía torácica
Digestología	Cirugía vascular
Neumología	Traumatología y cirugía ortopédica
Urología	Cirugía pediátrica
Neurología	ORL
Pediatría	Urología
Oncología	Oftalmología
Atención sociosanitaria y geriatría	Ginecología y obstetricia

Aparte de los servicios de soporte claramente identificados, en los procesos del hospital hay otros relacionados con la atención hospitalaria, como son:

- Las unidades de cuidados intensivos (UCI).
- Los servicios de anestesia y reanimación, también a menudo relacionados con áreas médicas para curas paliativas (la clínica del dolor).

Para simplificar nuestro caso, consideraremos solo la hospitalización convencional, excluyendo la actividad vinculada a la cirugía sin ingreso, la cirugía mayor ambulatoria o el hospital de día.

En lo que se refiere a las áreas de hospitalización, podemos considerar dos líneas bien diferenciadas de acceso:

- La admisión urgente, generalmente derivada de la actividad del servicio de urgencias.
- La admisión programada, que normalmente vendrá derivada de atención primaria o de la atención especializada (consulta externa) del propio centro.

Esta actividad programada tiene que ser gestionada por las áreas de gestión asistencial (gestión administrativa) y por los propios servicios clínicos, que deben priorizar y asignar sus recursos (generalmente basados en ocupación de camas o quirófanos) según la urgencia, criticidad y complejidad de cada caso, y según la disponibilidad de recursos. Esta gestión es lo que se conoce como gestión de la lista de espera.

Aparte de la gestión propia de la lista de espera, determinada por el propio servicio y, por extensión, del propio centro, dependiendo de su funcionamiento y de la disponibilidad de recursos, las autoridades sanitarias establecen objetivos que cumplir por parte de cada uno de los centros en sus contratos de compra de servicios, y que afectan al pago de los servicios adquiridos de acuerdo con los niveles de cumplimiento mínimos establecidos.

Cabe recordar que en el modelo de compra y facturación de los servicios establecidos por Catalunya, no solo se compra una determinada actividad quirúrgica sino que esta también debe prestarse de acuerdo con unos determinados indicadores de calidad (normalmente de tipo temporal). Los centros deberán garantizar el cumplimiento de estos indicadores para asegurar que se satisfacen adecuadamente los contratos de compra firmados con el Catalunya.

1.4. Enfermedades infecciosas

Actualmente existen multitud de enfermedades infecciosas que son causadas por diferentes agentes patógenos. En función del origen de la infección estas pueden ser clasificadas en:

- Infecciones bacterianas como el cólera, la neumonía, la tuberculosis el tétanos o la difteria.
- Infecciones víricas como la gripe, la hepatitis, el herpes, la mononucleosis, el dengue, la fiebre amarilla, el sarampión o la varicela.
- Infecciones por protozoos como la malaria o la amebiasis.
- Infecciones causadas por hongos.

Normalmente, las enfermedades infecciosas pasan por distintas etapas:

- Incubación: tiempo comprendido entre el comienzo de la infección hasta la aparición de sus síntomas. El tiempo de incubación varía en función de la enfermedad.
- Prodromia: consiste en el tiempo en que el cuerpo comienza a reaccionar a la infección. Este periodo acostumbra a ser muy corto y suele implicar una

sensación de malestar general, ligera elevación de temperatura, dolores musculares y en ocasiones náuseas y vómitos.

- Clínico: etapa donde aparecen los síntomas característicos de la enfermedad. En esta etapa la reacción del organismo ante la enfermedad es aparente.
- Tiempo de convalecencia: periodo de tiempo en el cual los síntomas de la enfermedad comienzan a desaparecer. Durante esta etapa el organismo elimina a los agentes de la enfermedad y se reparan los deterioros sufridos por esta.
- Recuperación: la evidencia de la enfermedad desaparece pero aún hay riesgo de contagio.

La gran mayoría de las enfermedades infecciosas no están presentes en países desarrollados o bien existe una vacuna preventiva para ellas. Sin embargo, existe todavía un grupo de enfermedades contagiosas para las que no se dispone de vacuna, como por ejemplo la gripe, el resfriado común, el virus del sida, el ébola, la polio y el lupus.

Algunos de los patógenos que causan estas enfermedades pueden llegar a nuestro territorio a través de viajeros o la importación de productos. Como consecuencia de ello, algunas enfermedades que se creían erradicadas en algunos países, como la sarna, han sufrido nuevos brotes. Aunque por suerte hasta la fecha han sido siempre controlados.

Entre las enfermedades infecciosas más comunes y con más impacto en la población en países desarrollados encontramos la gripe. Esta enfermedad afecta a las vías respiratorias y sus síntomas iniciales son similares a los de un resfriado común:

- Dolor, especialmente en las articulaciones (rodillas, codos y hombros), y en la musculatura de la espalda.
- Dolor en la faringe con sensación de sequedad en la garganta.
- Tos generalmente seca.
- Congestión nasal con estornudos y mucosidad transparente.
- Fiebre con escalofríos. La fiebre acostumbra a estar entre los 38 y los 41 grados centígrados. Suele durar entre uno y siete días, siendo 3 días lo habitual.
- Dolor de cabeza como consecuencia de la fiebre.

- Es habitual la secreción de lágrimas.
- Dolor ocular que se produce en el afectado cuando mira de forma lateral.
- Dolor en el abdomen, más habitual en niños.

La gripe se transmite desde individuos infectados a través de partículas en el aire cargadas de virus que son emitidas con la tos, los estornudos o al hablar. También es transmisible por la sangre y por las superficies u objetos contaminados con el virus.

Los virus de la gripe resisten más en ambientes fríos y secos y pueden conservar su capacidad infectiva durante días: durante una semana en temperaturas similares a las del cuerpo humano y durante más tiempo a medida que desciende la temperatura.

Para reducir los casos de gripe, en los países desarrollados se han establecido campañas de vacunación anual para las personas que tienen mayor riesgo de contraer la enfermedad o que son más vulnerables a sus complicaciones. La efectividad de las vacunas se reduce con el tiempo debido a las frecuentes y rápidas mutaciones que sufre el virus.

A pesar de las campañas de vacunación, con la atenuante de que la vacuna no es obligatoria y parte de la población no se vacuna, decenas de miles de personas en Cataluña (y cientos de miles en España) contraen el virus de la gripe cada año. La mayoría de los casos de infección ocurren en periodo invernal, cuando el virus es más resistente. Por otro lado predecir con exactitud la fecha exacta del auge del virus es harto complejo, y normalmente se descubre a posteriori, una vez ya se tiene evidencia de un gran número nuevo de contagios.

El virus de la gripe es un grave problema de salud pública, ya que ocupa parte de los recursos sanitarios durante un periodo de tiempo considerable y tiene importantes repercusiones económicas, no solo por los costes sanitarios que provoca, sino debido a la reducción de la productividad laboral de los afectados. De aquí la necesidad de utilizar nuevas técnicas capaces de predecir con mayor antelación la aparición de un nuevo brote.

1.5. Redes sociales

Internet y las redes sociales han revolucionado la comunicación entre personas. Gracias a las nuevas tecnologías, cualquier persona puede generar y compartir información con un gran número de personas en un tiempo récord. Además el volumen de información disponible es inmenso debido a que las redes sociales se han democratizado y han provocado un cambio cultural donde la gente está ávida por compartir información, tanto personal como profesional.

Gracias a la información obtenida de redes sociales, podemos realizar algunas tareas con mayor precisión: segmentar el mercado, mejorar la atención al cliente, estudiar la competencia, realizar estudios de mercado, anticiparnos a hechos que pasarán en el corto plazo, etc. Es esta última característica, la capacidad de anticiparse a los acontecimientos o de ser capaz de predecir un hecho con antelación la que permitiría predecir brotes víricos. Para hacerlo podríamos utilizar cualquiera de las redes sociales utilizadas por los internautas para compartir información personal (Twitter o Facebook) y/o para buscar información (Google). Debido al contexto de este caso práctico, de ahora en adelante, nos centraremos en el caso de Twitter.

Algunos sectores de la economía ya utilizan Twitter para predecir la evolución de su sector en el corto plazo. Por ejemplo, en el contexto de los mercados bursátiles, se utiliza Twitter para medir los cambios en la opinión pública sobre una acción o una empresa. Se ha observado que estas variaciones en el estado de ánimo acostumbra a ir acompañadas de una variación en la cotización de las acciones de la empresa involucrada en un plazo muy corto de tiempo. En el mismo contexto, Twitter también se utiliza para contar el número de mensajes escritos sobre una empresa y el incremento respecto al día anterior o respecto a la media de mensajes escritos sobre ella. Esta información ha resultado útil para predecir la cotización de la empresa en el cierre del día.

Sobre el uso de Twitter para medir los cambios en la opinión pública sobre las acciones o empresas, podéis leer el artículo titulado “Twitter predicts future of stocks”.

Debido a la gran cantidad de información disponible, el problema principal con el que se encuentran las empresas cuando analizan las redes sociales es el de seleccionar únicamente aquella información que es relevante para la toma de decisiones. Para ello, se pueden utilizar diferentes KPI en función de la red social con la que se trabaje y el objetivo perseguido.

Por ejemplo, en Twitter se pueden definir los siguientes KPI para medir los usuarios:

- Total seguidores.
- Nuevos seguidores de perfil (diarios, semanales, mensuales).
- *Engagement*¹:
 - Menciones.
 - Respuestas.
 - Retuits.
- Favoritos.

⁽¹⁾El *engagement* mide el grado de interrelación y fidelidad de un usuario con una marca, producto o servicio ofrecido por un tercero.

- Impresiones de un mensaje.
- Clics en los enlaces.
- Listas en las que una cuenta aparece.

Cada red social genera información diferente en función de cómo interactúan entre sí sus usuarios. A la hora de analizar estos datos, se deben estudiar los datos ofrecidos por las redes sociales para definir los KPI que permitan obtener información útil para la toma de decisiones. Estos indicadores pueden ser los KPI que ofrece la red social o indicadores calculados a partir de los KPI de la red social, como por ejemplo, la variación porcentual en un periodo de tiempo en el número de mensajes que hablan sobre un mismo tema/empresa.

2. Motivación

Actualmente, un gran número de agentes del sistema sanitario catalán tiene presencia en las redes sociales: ICS central (Twitter y Facebook), el hospital de Vall d'Hebron (Twitter y Facebook), el servicio de Atención Primaria de Badalona (Twitter), el Ayuntamiento de Barcelona mediante Salud Pública de Barcelona (Twitter), y un largo etcétera.

El uso de las redes sociales por los agentes está principalmente enfocado a informar a la ciudadanía. En particular, se informa de las novedades de los centros, de las investigaciones realizadas, de congresos y seminarios recientes, etc. Pero también tienen como objetivo obtener información sobre el grado de satisfacción de los pacientes.

No obstante, aplicado al sistema sanitario catalán, las redes sociales pueden tener muchas utilidades, como por ejemplo ayudar a predecir un incremento de personas afectadas por enfermedades infecciosas, lo que permite informar mejor a la población con carácter preventivo y gestionar de forma más eficiente los recursos sanitarios para atender a los afectados.

En los periodos de auge de una enfermedad vírica, el número de búsquedas y comentarios realizados por personas sobre la enfermedad aumenta de forma muy significativa; hay un incremento inicial moderado en una etapa inicial o prevírica y un gran aumento cuando el virus se encuentra en su periodo de auge. Gracias a esta información, Google ha lanzado el proyecto Google Flu-Trends, que mediante el recuento de consultas realizadas sobre la gripe estima en qué medida circula la gripe por diferentes países y regiones del mundo.

Dejando de lado lo que puede conseguirse mediante el análisis de búsquedas en Google, Twitter permite recuperar aquellos mensajes que contengan las palabras clave de la búsqueda en el *hashtag*, el nombre del usuario o el contenido del mensaje, lo que permite recuperar solo aquellos mensajes emitidos desde una zona geográfica y en un intervalo de tiempo. Esto permitiría recuperar los tuits de Cataluña que traten sobre la gripe. Un análisis del número de tuits (comparándolo con número de tuits en el pasado) podría permitir estimar si nos encontramos en una situación similar a la acontecida otros años antes de la aparición de un brote vírico.

Hashtag

Un *hashtag* (o etiqueta) es una cadena de caracteres formada por una o varias palabras concatenadas y precedidas por una almohadilla (#). Es, por lo tanto, una etiqueta de metadatos con el fin de que tanto el sistema como el usuario identifiquen el contenido del mensaje de forma rápida.

3. Enunciado del caso práctico

El objetivo de este caso práctico es diseñar un sistema capaz de prever el aumento de casos de gripe a partir de la información recuperada de Twitter y de los datos de urgencias de los centros médicos. El sistema deberá predecir el incremento de los casos de gripe antes de que estos se produzcan y permitir así una mejor gestión de sus recursos a las autoridades sanitarias.

3.1. Trabajo que realizar

Tal y como se ha comentado, el objetivo de este caso práctico es generar un sistema predictivo que permita estimar brotes gripales antes de que ocurran y integrar este sistema a los sistemas de inteligencia de negocio de la empresa.

Para ello dividiremos el trabajo en 4 partes:

1) Recuperación de la información

Recuperación de los mensajes de Twitter susceptibles de hablar de la enfermedad. Con el fin de poder recuperar estos mensajes, tendremos que definir las palabras clave que deben contener los mensajes: gripe, gripe A, gripe B, virus, etc. Se debe tener en cuenta que búsquedas con palabras clave como “gripe” y “gripee” devolverán resultados diferentes.

Las búsquedas deberán estar geolocalizadas, y se recuperarán solo aquellos mensajes enviados dentro del área de actuación del ICS o de regiones vecinas que puedan tener influencia en la propagación del virus de la gripe.

Asimismo, se deberá tener en cuenta el idioma de los mensajes para definir las cadenas de búsqueda, ya que muchos usuarios escriben mensajes tanto en su idioma materno como en inglés.

2) Creación de un cuadro de mando

La segunda parte consistirá en la creación de un cuadro de mando con un sistema de indicadores que permita comparar el impacto actual de la gripe en Twitter respecto al impacto en Twitter en otros periodos pasados que hayan precedido a un brote del virus de la gripe.

El estudiante deberá resolver detalladamente cada uno de los pasos desde el planteamiento estratégico a la definición de bajo nivel de los indicadores y KPI que se definan. Para ello, el estudiante deberá pensar en posibles KPI secundarios, elaborados a partir de los datos recogidos en las redes sociales, para

crear indicadores de alerta que permitan verificar a los diferentes agentes del sistema sanitario si nos encontramos en una situación que preceda a un brote viral, y hasta qué punto el brote viral está en sus cercanías geográficas.

A modo de ejemplo, algunos KPI secundarios podrían ser:

- Incremento porcentual en el número de mensajes semanalmente.
- Similitud en el número de mensajes en Twitter entre la semana actual y las épocas precedentes a un brote.
- Una combinación de los dos indicadores anteriores.

Los KPI anteriores podrían presentarse de forma geográfica por provincias.

Para poder contrastar sus indicadores numéricamente, el usuario podrá hacer uso de las fuentes de datos disponibles de hospitalización, urgencias y mensajes de Twitter para encontrar una relación entre los mensajes publicados en Twitter, las visitas a urgencias y las hospitalizaciones por casos de gripe.

Nota

Junto al enunciado se proporcionan 2 ficheros excels con datos sobre la hospitalización, urgencias y mensajes de Twitter.

3) Carga de datos

Se pide el desarrollo de los procesos de carga de la información recuperada de las redes sociales.

El estudiante deberá diseñar un almacén de datos donde almacenar los datos recuperados de Twitter. Opcionalmente, el estudiante podrá integrar dicha información en el almacén de datos creado en la asignatura de *Data Warehouse*. Sea cual sea la elección del estudiante, se deberá razonar si los KPI definidos en la segunda parte del ejercicio se incluyen en el *datawarehouse* existente o en uno de nueva creación.

A partir del análisis de las fuentes de datos y de las características del API de Twitter, el estudiante deberá diseñar el proceso de recuperación de nueva información de la red social y su almacenamiento en el datawarehouse escogido, teniendo en cuenta la periodicidad en la recuperación de datos y la posible aplicación de un filtrado de contenido en los mensajes. Se recomienda estudiar a fondo la sección 3.3 (“Notas técnicas: recuperación de datos de Twitter”) antes de abordar este diseño.

4) Explotación de los datos

Por último, el estudiante deberá indicar cómo el sistema responde a las preguntas mínimas de los usuarios y plantear posibles acciones futuras, como la comparación con nuevos KPI o KPI descartados en la segunda parte del caso para aumentar la efectividad del sistema.

3.2. Fuentes de datos suministradas

Para el presente caso se proporcionan datos de hospitalización, urgencias y tuits de usuarios. Estos ficheros son el resultado del funcionamiento operativo de un hospital básico y recogen la información principal en términos de paciente, procedimiento, fechas, tratamiento, etcétera y los mensajes sobre gripe escritos por usuarios de Twitter.

Los mensajes de Twitter se han extraído mediante consultas a la API. En particular responden a los tuits realizados desde Cataluña que hablan sobre la gripe. Para cada tuit se muestra su contenido y la fecha en que fue escrito.

Los ficheros han sido tratados para que cumplan las condiciones de la Ley orgánica de protección de datos (LOPD) y en ningún momento se vulneren los derechos de los pacientes.

3.3. Notas técnicas: Recuperación de datos de Twitter

Las redes sociales ofrecen API para facilitar la búsqueda, filtraje y recuperación de datos.

Para recuperar los datos se pueden utilizar bibliotecas que interactúen con la API de Twitter, ayudarse de alguna herramienta web o utilizar la consola de Twitter para desarrolladores. Para una primera exploración se puede utilizar la opción de búsqueda avanzada de la interfaz web de Twitter, aunque las opciones que permite personalizar la búsqueda son limitadas.

Antes de automatizar la recuperación de resultados con su API u otros mecanismos, Twitter recomienda hacer una verificación manual siguiendo los siguientes pasos:

1) Realizar la búsqueda deseada en <http://twitter.com/search> y copiar la URL resultante de la búsqueda. Por ejemplo, una búsqueda de los tuits que contengan la palabra gripe daría como resultado la siguiente URL: <https://twitter.com/search?q=gripe&src=typd>.

2) Substituir <https://twitter.com/search> por <https://api.twitter.com/1.1/search/tweets.json> (en nuestro ejemplo la URL resultante sería: <https://api.twitter.com/1.1/search/tweets.json?q=gripe&src=typd>). Haciendo esto creamos una llamada a la API de Twitter donde pedimos que se devuelvan los tuits que tengan la palabra “gripe” en formato JSON.

3) Ejecutar la URL obtenida en el segundo paso en la consola de desarrolladores de Twitter. Para ejecutar la URL tendremos que escoger la opción de autenticación “Oauth 1” e introducir los datos de nuestra cuenta de Twitter. Después de ejecutar la llamada podremos comprobar el resultado en formato JSON.

API

Una API es una interfaz de programación de aplicaciones (en inglés, *application programming interface*).

Herramientas web

Hay muchas webs que permiten recuperar y analizar datos de Twitter y estas están en continua evolución. En el enlace tenéis una lista de algunas de ellas.

Verificación manual

Para verificar una búsqueda antes de automatizarla con la API, los pasos que seguir se explican en este enlace.

Como podéis ver en la URL que ha generado Twitter (<https://twitter.com/search?q=gripe&src=typd>), el parámetro *q* se utiliza para indicar la palabra que hay que buscar. Twitter permite utilizar otros parámetros en la búsqueda. A continuación se muestran los más relevantes para esta actividad.

Tabla 2. Parámetros de búsqueda más relevantes

Nombre del parámetro	Significado
<i>Q</i> (obligatorio)	La cadena de texto que se busca
<i>Geocode</i>	Devuelve solo los tuits de los usuarios localizados en un radio <i>x</i> de las coordenadas geográficas indicadas
<i>lang</i>	Devuelve solo los tuits en el idioma indicado
<i>locale</i>	Indica el idioma utilizado en la cadena de texto que se busca
<i>result_type</i>	Indica el tipo de resultados que se desean recibir. El valor por defecto es <i>mixed</i> y las posibilidades son: <ul style="list-style-type: none"> • <i>recent</i>: los resultados más recientes • <i>popular</i>: los resultados más populares • <i>mixed</i>: tuits populares y recientes
<i>Until</i>	Tuits generados hasta la fecha indicada
<i>Since</i>	Tuits generados desde la fecha indicada

Parámetros de búsqueda

Podéis ver todos los parámetros para personalizar las búsquedas en este enlace.

Idiomas

En este enlace se pueden encontrar los idiomas con los que se puede personalizar la búsqueda en Twitter. "en" para inglés, "es" para español, "fr" para francés, etc. En catalán no está disponible.

Por ejemplo, si quisiéramos recuperar mensajes que contengan la palabra gripe introducidos en el área de Barcelona, podríamos utilizar la siguiente consulta:

```
https://api.twitter.com/1.1/search/tweets.json?
q=gripe&src=typd&geocode="41.4166,2.1846,50km"
```

Notad que se usa el parámetro *q* para indicar que estamos buscando tuits que contengan la palabra gripe y se usa el parámetro *geocode* para indicar que los tuits deben haber sido hechos en un área de 50 km de diámetro centrada en las coordenadas 41.4166 N, 2.1846 E. Podéis ver dichas coordenadas localizadas en el mapa preguntando a un servicio de mapas, como Google Maps o OpenStreetMaps, sobre estas coordenadas 41.4166, 2.1846.

Por otro lado, si queremos recuperar aquellos mensajes en inglés escritos en Barcelona y alrededores, añadiríamos el parámetro *lang* con valor "en", que es el valor del idioma inglés:

```
https://api.twitter.com/1.1/search/tweets.json?
q=gripe&src=typd&geocode="41.4166,2.1846,50km"&lang=en
```

Cuando se realicen consultas usando la API de una red social se deberá tener en cuenta el número máximo de peticiones que se pueden realizar y el intervalo de espera entre peticiones. Esto es necesario porque el número de resultados devueltos en cada petición está acotado y si no recuperamos la informa-

ción frecuentemente podemos perder una gran cantidad de datos. De hecho, si excedemos, en un intervalo de tiempo, el número máximo de peticiones que Twitter permite, nuestra aplicación de recuperación será bloqueada por Twitter y esta no podrá realizar más consultas hasta pasado un tiempo.

Número máximo de consultas

En 2014, el número máximo de resultados devueltos en cada consulta es de 100. El número máximo de consultas que pueden realizar las aplicaciones cada 15 minutos es de 450. Si supera ese límite la aplicación será bloqueada. Es importante saber que estos valores cambian constantemente, por tanto, es importante comprobar los límites antes de realizar un proceso de extracción de datos.

