

## 5 Business intelligence y DSS

### 5.1 Introducción

Como ya vimos en el primer tema, las empresas han desarrollado los sistemas de información operacionales, también conocidos como sistemas transaccionales, para poder dar soporte y automatizar sus procesos de negocio. Estos sistemas están orientados a la captura rápida y eficaz de los datos correspondientes a los distintos eventos de negocio (transacciones). Sin embargo, estos sistemas han prestado un escaso soporte al proceso de toma de decisiones, el cual, en muchos casos, requiere explotar la información generada a partir de los datos acumulados en el desarrollo de las actividades de su negocio.

Los sistemas transaccionales surgen con las primeras aplicaciones empresariales de los equipos informáticos, para realizar de forma automatizada tareas administrativas repetitivas e intensivas en mano de obra: la gestión de nóminas, la facturación a clientes, el control de inventarios, la contabilidad, etc.

Sin embargo, los sistemas informacionales utilizan los datos almacenados en los sistemas informáticos de la empresa (capturados y almacenados por el sistema transaccional) para generar información útil para el proceso de toma de decisiones. Así mismo, estos sistemas pueden incorporar datos procedentes de fuentes externas para su análisis.

Los sistemas transaccionales están orientados al registro de los eventos de negocio, con todo el nivel de detalle, facilitando la realización de operaciones frecuentes de actualización, inserción, consulta y eliminación de datos. Por este motivo, en estos sistemas se hace especial énfasis en garantizar la consistencia de los datos y su seguridad, minimizando los tiempos de respuesta (no sería aceptable que el usuario del sistema tuviera que esperar varios minutos para registrar un nuevo pedido de un cliente).

Los sistemas transaccionales pueden proporcionar alguna información básica para la gestión, a partir de consultas e informes predefinidos, pero generalmente prestan poca atención a los datos acumulados sobre el negocio, debido sobre todo a razones de eficiencia. Sus usuarios dentro de una organización tienen unos perfiles medios o bajos en cuanto a su responsabilidad y capacidad para tomar decisiones.

Por su parte, los sistemas informacionales están orientados al análisis de los datos acumulados y a la simulación de alternativas, como soporte al proceso de toma de decisiones. Por ello, están especializados en la consulta y no en la actualización, trabajan con grandes volúmenes de datos no volátiles (estos datos "son historia", se corresponden con transacciones ya completadas y que, por lo tanto, no se tienen que actualizar).

Sus usuarios tienen unos perfiles altos en cuanto a su responsabilidad y capacidad para tomar decisiones, ya que en muchos casos la información obtenida se va a utilizar para el análisis y planificación estratégica dentro de la organización.

Por lo tanto, en los sistemas informacionales se parte de los datos acumulados por el negocio para poder llevar a cabo un análisis de ellos mediante consultas y procesos masivos, con una proyección de los resultados hacia el presente y el futuro.

## 5.2 Estructura de un sistema de business intelligence

Según Gartner Group, se entiende como herramientas de *business intelligence* todas las aplicaciones orientadas a convertir datos en conocimiento para la organización, mediante la transformación de datos en información estructurada para su explotación directa por parte de la dirección.

Estas herramientas incluyen un conjunto muy amplio de aplicaciones orientadas a los siguientes procesos:

- Extracción de datos.
- Almacenamiento.
- Modelado y simulación.
- Explotación de la información, gestión de alertas.
- Modelos de gestión: financiero, comercial, presupuestario, etc.

Desde un punto de vista conceptual, herramientas como la hoja de cálculo Excel podrían cubrir los aspectos anteriores, sobre todo si se utilizan sus funciones más avanzadas de tratamiento de datos, como las tablas dinámicas. No obstante, en este mercado se ha desarrollado un conjunto de productos orientados a uno o más de los procesos anteriores, añadiendo posibilidades entre las que destaca el poder manejar grandes volúmenes de datos, superando con creces las limitaciones de una hoja de cálculo en este aspecto. Tal como apunta el término "*business intelligence*", este tipo de herramientas se apoya en muchos casos en técnicas propias del área de la inteligencia artificial y en técnicas y herramientas afines.

En los últimos años se ha producido una entrada en el mercado de soluciones *business intelligence* de los fabricantes de productos ERP como estrategia para añadir mayor valor a su oferta. La gran ventaja es que los proveedores de ERP conocen sus modelos de datos y pueden ofrecer soluciones integradas que exploten datos procedentes del ERP en las distintas áreas (financiera, comercial, producción). Inicialmente algunos de estos proveedores ofrecieron tecnología propia, pero en los últimos años hemos asistido a la fusión y compra de productos de *business intelligence* por parte de los fabricantes de productos ERP, e incluso a la adquisición de empresas dentro del propio mercado *business intelligence* para reforzar las capacidades de los productos por la vía de la adquisición de la tecnología.

En la figura 5.1 se muestra la estructura arquitectónica de una solución de *business intelligence* en la que las distintas capas de software ofrecen los distintos niveles de servicio necesarios para que puedan funcionar los sistemas de apoyo a la toma de decisiones (*decision support systems* - DSS). En la base tenemos los sistemas transaccionales, que son los que proporcionan la materia bruta (datos) que se almacenan de una forma más racional mediante sistemas de almacenes de datos (*data warehousing*) en contraste con los sistemas de bases de datos tradicionales. El disponer de los datos en un sistema de almacenamiento de datos facilita la labor de las distintas técnicas de análisis de datos, de minería de datos y de inteligencia artificial que permiten obtener el conocimiento necesario para que las herramientas DSS puedan prestar los servicios que requieren los directivos.

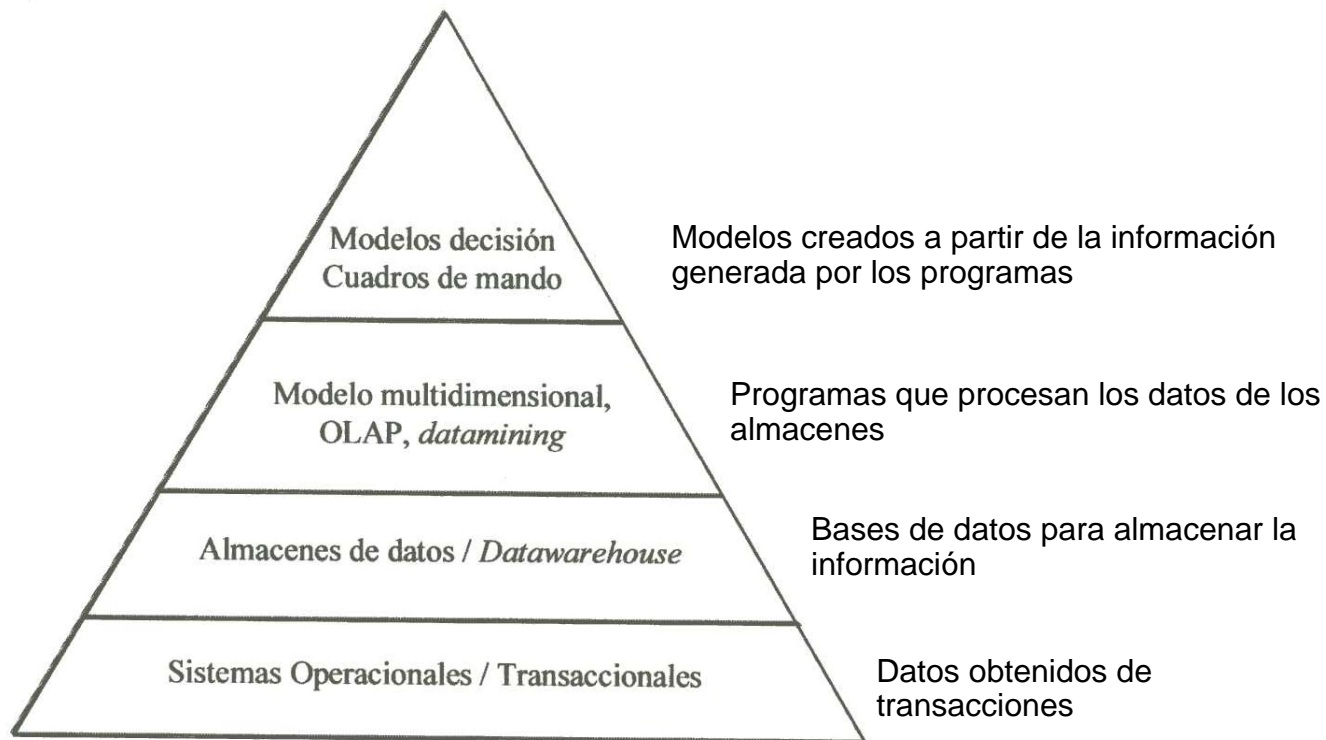


Figura 5.1. *Jerarquía de soluciones de business intelligence*

Las fases necesarias para implantar una solución basada en herramientas de *business intelligence* son las siguientes:

- **Definición de las necesidades de información y del origen de los datos.** En esta primera fase se deben analizar y definir de forma detallada las especificaciones para los distintos usuarios del sistema, estableciendo además el origen de los datos, que puede ser múltiple. Es importante destacar que este tipo de sistemas suelen requerir la integración de datos procedentes de distintas fuentes, incluyendo datos de procedencia externa de la organización.
- **Diseño de la arquitectura tecnológica:** herramientas de *business intelligence* y plataformas hardware y software. En esta fase se deben definir los productos a integrar y las necesidades desde el punto de vista de almacenamiento y proceso necesarias.
- **Construcción del *data warehouse* y de los modelos de explotación de los datos,** que incluye el desarrollo de los programas de carga de datos y también de todas las aplicaciones de visualización.

Conviene tener en cuenta que los sistemas de *data warehousing* y de minería de datos requieren equipos de altas prestaciones para poder manejar grandes volúmenes de datos con rapidez y eficacia. Por este motivo, se suelen emplear servidores multiprocesador, con plataformas fácilmente escalables, que permitan soportar crecimientos en el volumen de datos procesados.

En las siguientes secciones se describen los componentes más especializados de esta arquitectura, omitiéndose los más básicos, es decir, los sistemas transaccionales.

### 5.3 Almacenes de datos

Un almacén de datos (*data warehouse*) constituye un elemento fundamental para un sistema informacional de apoyo a la toma de decisiones, ya que es precisamente el componente que permite almacenar la información.

En un almacén de datos se almacenan datos procedentes de varias fuentes: datos procedentes de los sistemas transaccionales y datos procedentes de fuentes externas. Todos estos datos, independientemente de su origen, son datos no volátiles, es decir, que son datos que no van a modificarse.

Un almacén de datos debe disponer de una gran capacidad de almacenamiento (de varias decenas de gigabytes, llegando incluso a alcanzar los terabytes y aun dimensiones mayores en las grandes empresas), ya que los datos pertenecen a largos periodos de tiempo (varios ejercicios económicos de la empresa).

El concepto de "metadatos" hace referencia a los datos que se almacenan acerca de los propios datos, lo que conocemos como diccionario de datos o catálogo del sistema, en el que figuran las definiciones de los atributos de datos almacenados y su estructura, pero también puede almacenarse otra información sobre los datos, como las transformaciones que han sufrido, su origen, etc.

En un almacén de datos se cumple un principio arquitectónico fundamental: la separación de los sistemas transaccionales de los informacionales en dos entornos tecnológicos diferentes, para que el análisis de los datos acumulados del negocio no interfiera con el procesamiento y registro de nuevas transacciones.

Estos datos se guardan de forma unificada, homogénea y accesible, con distintos niveles de agrupación. Asimismo, los datos en un almacén de datos se encuentran organizados por temas (clientes, vendedores...) y no por funciones, como en los sistemas transaccionales, es decir, todos los datos relativos a los clientes se encuentran dentro de la misma estructura de datos en el almacén, mientras que en un sistema transaccional se reparten entre distintas funciones: marketing, ventas, contabilidad, post-venta, etc.

En estos sistemas también se utiliza el concepto de *data mart* para referirse a un almacén de datos de menor tamaño, restringido generalmente a un área temática o a un departamento de la empresa.

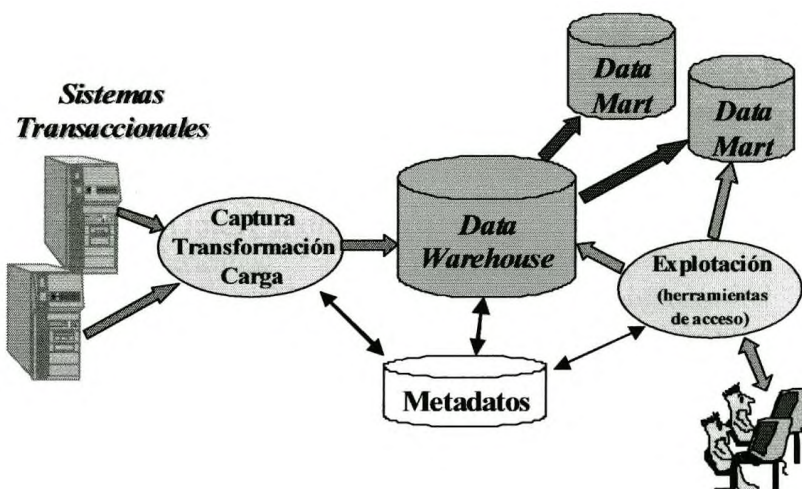


Figura 5.2. Estructura de un sistema de data warehouse

### 5.3.1 Etapas en la construcción de un almacén de datos

#### 5.3.1.1 Captura de datos

En esta primera etapa se utilizan herramientas de **extracción que soportan múltiples formatos de almacenamiento de los datos** (distintos formatos de bases de datos, tablas de texto, hojas de cálculo, etc.), para incorporar los datos que han sido seleccionados al almacén.

#### 5.3.1.2 Tratamiento, conversión y transformación de los datos

Una vez cargados los datos, es necesario realizar una serie de **operaciones de limpieza, homogeneización, mezclado y enriquecimiento de los datos**, entre las que podemos destacar las siguientes:

- Detectar y corregir errores.
  - Eliminar registros duplicados (por ejemplo, clientes dados de alta más de una vez).
  - Detectar y anular valores sin sentido (por ejemplo, fechas de alta anteriores al comienzo de la actividad).
- Analizar la consistencia en el uso de los valores. Dado que los datos cargados en el sistema provienen de distintas fuentes y departamentos, es muy probable que se hayan utilizado distintos esquemas de codificación de valores, por lo que se hace necesario llevar a cabo un proceso de homogeneización:
  - Codificación del sexo: varón / hembra, hombre / mujer, 1 / 0, etc.
  - Utilización de distintas unidades de medida: metros, centímetros, pulgadas...
  - Formato de las fechas: dd/mm/aaaa, mm/dd/aaaa, etc.
- Tratamiento de la valores nulos.
  - Asignación de valores por defecto.
- Eliminar campos no significativos. Eliminar campos que no van a aportar nada al análisis posterior como, por ejemplo, aquellos utilizados en el sistema transaccional para reflejar el estado de una transacción (todas las transacciones han sido completadas ya cuando se incorporan al almacén).
- Reestructurar y añadir nuevos campos. Enriquecimiento de los datos, recabando, si es posible, más información sobre las entidades objeto de análisis (clientes, proveedores, productos, etc.).
- Codificar campos para facilitar su tratamiento estadístico.
  - Pasar de direcciones a códigos de zonas geográficas.
  - Pasar de fechas de nacimiento a intervalos de edades.
  - Pasar de atributos booleanos a valores binarios (0 ó 1).
- Cálculo de campos derivados. Cálculo de subtotales y de datos consolidados que reduzcan el nivel de detalle: ventas por región, ventas por producto, etc.

### 5.3.2 Sistema gestor de datos en un almacén de datos

En un almacén es posible utilizar tres tipos de sistemas de gestión de bases de datos:

Base de datos relacional tradicional.

Base de datos relacional con un diseño en estrella y una desnormalización de tablas.

Base de datos multidimensional.

#### 5.3.2.1 Base de datos relacional tradicional

Podemos utilizar una base de datos relacional convencional (Access, SQL Server, Oracle...) para construir sobre ella un almacén, sobre todo si éste es de un tamaño relativamente pequeño.

Sin embargo, debemos tener en cuenta que en un almacén deben reconsiderarse los principios de diseño de las bases de datos relacionales. Un sistema relacional se ha diseñado para poder gestionar un número elevado de transacciones por segundo, pero en cada transacción, frecuentemente, se pretende acceder a un número de datos pequeño (recuperación de un registro, actualización de un dato...).

Asimismo, en los sistemas de gestión de bases de datos relacionales se presta especial atención a la consistencia y la integridad de los datos y, por este motivo, se lleva a cabo durante la etapa de diseño de las estructuras de datos un proceso de normalización. De este modo, se consiguen eliminar redundancias (los atributos no se repiten en distintas tablas de datos).

Pero debido a este proceso de normalización, las tablas se suelen descomponer en un mayor número de tablas más pequeñas (con menos columnas), lo que implica que las consultas muchas veces deban reunir varias tablas para ejecutarse, siendo la operación de reunión una operación costosa.

Hay que tener en cuenta que en un almacén se gestiona un número muy inferior de operaciones, pero en cada una de ellas se han de manipular una gran cantidad de datos y, además, estos datos no pueden ser modificados por los usuarios (son datos no volátiles), por lo que ya no tiene tanta importancia llevar a cabo un proceso de normalización, sobre todo teniendo en cuenta que el coste de almacenamiento de la información ha descendido drásticamente en los últimos años.

#### 5.3.2.2 Base de datos relacional con diseño en estrella

Esta variante en el diseño de las bases de datos relacionales consiste básicamente en utilizar estructuras de datos no normalizadas que incorporan redundancias que permiten agilizar las consultas y operaciones de análisis de datos.

#### 5.3.2.3 Bases de datos multidimensionales

Estas bases de datos han surgido recientemente como una alternativa para guardar los datos en un almacén de datos. En estas nuevas estructuras se almacenan los datos en "cubos multidimensionales", especialmente diseñados para acelerar las consultas y el análisis multidimensional de la información.



Por este motivo, se lleva a cabo un proceso de cálculo de subtotales durante la etapa de carga del sistema, con varios niveles de agrupamiento, asignando cada uno de estos subtotales a las distintas celdas que constituyen el cubo multidimensional, tal y como se refleja en la figura 5.3.

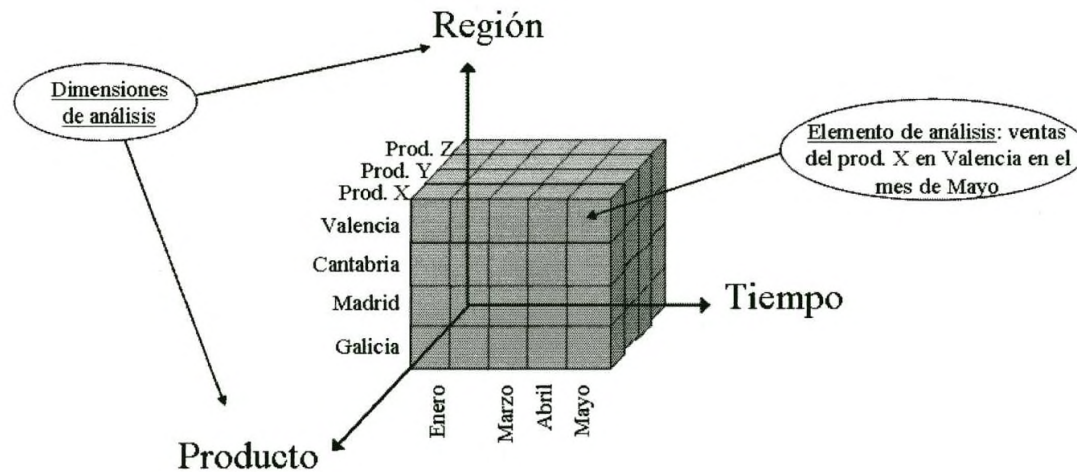


Figura 5.3. Base de datos multidimensional

No obstante, estas bases de datos también presentan varios inconvenientes: por una parte, tienen importantes limitaciones para su escalabilidad, ya que agregar nuevos datos al sistema obliga a recalcular todos los subtotales, es decir, a reconstruir toda la estructura de celdas del cubo multidimensional. Por otra parte, hay que tener en cuenta que estas estructuras ocupan un gran espacio de almacenamiento, espacio que crece de forma exponencial al considerar nuevas dimensiones de análisis.

En comparación con las bases de datos multidimensionales, las relacionales en estrella presentan tiempos de consulta mayores, ya que los datos no están precalculados. Sin embargo, consiguen reducir el tamaño de la base de datos y mejorar su escalabilidad, ya que al añadir nuevos datos al sistema no obliga a recalcular subtotales. Además, aportan una mayor flexibilidad en las consultas.

#### 5.4 Procesamiento analítico en línea (OLAP)

Las técnicas OLAP (*on-line analytical processing*) emplean un modelo multidimensional, constituido por tres componentes:

- **Dimensiones:** grupos conceptuales que permiten analizar o consolidar los datos, por ejemplo, productos, clientes, zonas geográficas, tiempo...
- **Medidas o indicadores:** valores numéricos que se guardan en la base de datos, por ejemplo, facturación, unidades vendidas...
- **Jerarquías de dimensiones:** distintos niveles de agregación dentro de una dimensión. Por ejemplo, dentro de la dimensión "zonas geográficas", se pueden definir las jerarquías de "ciudades", "provincias", "regiones", "países", etc.

Las operaciones básicas de análisis multidimensional permiten navegar por los datos almacenados en el almacén de datos. Seguidamente, se presentan algunas de estas operaciones básicas:

- Cambiar de dimensión de análisis (*drill across*).
- Permutar dos dimensiones de análisis (*swap*).
- Subir (*up*) o descender (*down*) en el nivel de agregación.
- Profundizar para alcanzar datos de un nivel inferior (*drill down*).
- Expandir un determinado nivel de información (*expand*).
- Anular la expansión de un nivel de información (*collapse*).

Los sistemas OLAP pueden estar apoyados por bases de datos relacionales (sistemas ROLAP: *relational OLAP*) o bien por cubos multidimensionales (sistemas MOLAP: *multidimensional OLAP*). También se han propuesto sistemas híbridos (sistemas HOLAP: *hybrid OLAP*), que combinan los dos tipos de estructuras de datos.

### 5.5 Minería de datos

Las herramientas de minería de datos (en inglés *data mining*, pero también *knowledge discovery in databases - KDD*) constituyen métodos avanzados para explorar y modelar relaciones en grandes volúmenes de datos y obtener información que se encuentra implícita: patrones de comportamiento de los clientes, asociaciones de productos, relación entre variables como el precio de un producto y su nivel de consumo, etc. Puede definirse la minería de datos como el proceso de búsqueda y análisis de datos para revelar información oculta y potencialmente valiosa para la organización.

La metodología seguida en un proceso de minería de datos puede constar de las siguientes etapas:

- **Muestreo:** selección de una muestra de datos, que permite reducir el coste del análisis e incrementar su velocidad. Se trata de una etapa aconsejable para volúmenes muy elevado de datos.
- **Exploración:** determinación de las tendencias principales, rango de las variables clave, frecuencia de los valores, correlación entre variables, etc.
- **Modificación:** transformación y filtrado de variables, para adecuarse a los requisitos del problema o cuestión que se pretende analizar.
- **Modelado del conocimiento:** en esta tarea se pueden emplear redes neuronales, árboles de decisión, técnicas de análisis estadístico multivariante, análisis de regresión, etc.
- **Evaluación:** comprobación de la validez del modelo obtenido.
- **Presentación gráfica de los resultados.**

~~Entre los distintos tipos de herramientas utilizados en los paquetes de minería de datos, podemos distinguir tres grandes grupos:~~

Entre las diferentes técnicas que suelen implementar las herramientas de minería de datos podemos destacar:



- **Herramientas estadísticas**
  - Cálculo de distintos parámetros estadísticos: medias, varianzas, correlaciones, etc.
  - Técnicas bayesianas.
  - Tests de hipótesis.
  - Técnicas de regresión lineal.
  - Análisis multivariante.
  - Análisis de agrupación por similitudes (*clustering*).
- **Técnicas de inteligencia artificial**, como por ejemplo:
  - **Algoritmos evolutivos**: una familia de técnicas de optimización y búsqueda que tienen como característica común su inspiración en los procesos evolutivos darwinianos, en las que las que se evoluciona un conjunto de soluciones candidatas a un problema y este conjunto de soluciones se somete a un proceso evolutivo a lo largo de una serie de generaciones durante el que dichas soluciones van evolucionando en base a procesos de mutación y cruce, para tratar de que sobreviva la solución más apta, es decir, aquella que es mejor para el problema. Dentro de esta familia de técnicas destacan los algoritmos genéticos y la programación genética.
  - **Redes neuronales**: una red neuronal trata de simular la capacidad de aprendizaje del cerebro humano, a partir de una serie de entradas de información y de un modelo compuesto por varias capas de decisión para la generación de una respuesta. Mediante el ajuste del modelo se consigue dar respuesta a nuevas situaciones. Para ello es necesario determinar el número de capas y de elementos de ponderación, así como proceder al entrenamiento de la red a partir de una muestra de datos inicial.

Mediante las técnicas de **inteligencia artificial se puede inferir conocimiento de distintas clases**, para las que se emplean diferentes formalismos de representación:

- **Árboles de decisión**: empleados en los algoritmos de clasificación.
- **Reglas de asociación**: que permiten establecer conexiones entre uno o más eventos, evaluando la correlación entre distintos tipos de información. Un ejemplo de regla de asociación podría ser similar a la siguiente: "el 30% de los compradores de un DVD poseen un reproductor de música MP3".
- **Patrones secuenciales**: extensión de la técnica de reglas de asociación que permite evaluar la ocurrencia de determinados eventos en una secuencia temporal. Un ejemplo de un patrón secuencial se podría formular de la siguiente forma: "el 50% de los nuevos clientes de telefonía móvil de tercera generación va a gastar más de 60 euros en servicios multimedia en los siguientes dos meses a la compra de su teléfono".

Entre las aplicaciones típicas de las técnicas de minería de datos, podemos citar las siguientes:

### Estudio de las asociaciones de productos y análisis de afinidades

Permiten analizar el nivel de correlación en el consumo de distintos productos (afinidad entre productos), así como su relación con los distintos tipos o grupos de clientes de la empresa (afinidad entre productos y clientes). De este modo, mediante el análisis de afinidades es posible determinar cuáles son los principales patrones de compra.

Así, por ejemplo, las técnicas de análisis de la cesta de la compra (*market basket analysis*) y de las asociaciones de productos dentro de ésta han puesto de manifiesto que una pareja de productos de gran venta en los supermercados de Estados Unidos son los pañales y la cerveza.

La información obtenida de este modo se puede utilizar en el diseño de catálogos de productos, en la organización de lineales en un almacén (se trataría en este caso de distanciar los productos que se suelen comprar conjuntamente, como los pañales y la cerveza, para que el consumidor tenga que recorrer una mayor distancia por el interior del supermercado, con lo que se incrementa la posibilidad de que incorpore más productos a su cesta de la compra), en la definición de políticas de promoción de ciertos productos, en el desarrollo de acciones comerciales y sistemas de recomendación para estimular la venta cruzada, etc.

### Segmentación y agrupamiento de clientes (*clustering*)

Permiten llevar a cabo la identificación de grupos de clientes con características sociodemográficas, hábitos y patrones de consumo similares. Con esta información es posible obtener una información más detallada de la composición de la cartera de clientes de la empresa, pudiendo especificar para cada segmento una ficha descriptiva como la siguiente:

- Características sociodemográficas básicas: edades, sexos, nivel de ingresos, idioma, etc.
- Características de distribución geográfica.
- Nivel actual de consumo de productos y estimación del nivel potencial.
- Hábitos de comportamiento y patrones de consumo.
- Rentabilidad obtenida.
- Nivel de vinculación y fidelización con la empresa.
- Adecuación de la oferta a las necesidades y características específicas del segmento.
- Sensibilidad a la variación de las condiciones comerciales actuales: precios, descuentos, envasado de los productos, etc.
- Grado de concentración del negocio en cada segmento e importancia para la empresa.

De este modo, se podría realizar una asignación de una probabilidad de compra de productos a los diferentes grupos, o bien tratar de obtener una "cartera de productos tipo" que cubra las necesidades de un determinado cliente por comparación con el grupo o segmento donde se encuadra.

Algunas empresas han desarrollado una matriz de productos y clientes para facilitar el análisis de la propensión de consumo de cada producto por parte de los distintos segmentos de clientes.

Otra posible aplicación sería el diseño de campañas comerciales y de programas de fidelización ajustados a los distintos grupos y segmentos de clientes identificados.

### **Clasificación de clientes**

En este caso se trataría de identificar el grupo al que pertenece un determinado cliente, información que podría ser utilizada en aplicaciones como las siguientes:

- Técnicas de puntuación (*scoring*), que tratan de determinar la importancia económica de cada cliente atendiendo al importe medio de sus compras, a la frecuencia de sus compras y a la fecha de la última compra, así como a otros datos de carácter socioeconómico. De este modo, es posible conocer la rentabilidad potencial de cada cliente y promover acciones comerciales con el objetivo de aumentar los ingresos medios por cliente, mediante estrategias de *up-selling*<sup>1</sup> y de venta cruzada.
- Análisis de riesgos en operaciones a crédito, minimizando el riesgo de impagados y de operaciones fallidas. Se trataría, por lo tanto, no solo de vender, sino de vender bien.
- Prevención y reducción del fraude.

### **Predicción del comportamiento de clientes**

El comportamiento inicial observado en los clientes puede ayudar a predecir su comportamiento futuro. Para ello, se podría llevar a cabo un análisis de la evolución en el tiempo del comportamiento de los clientes, empleando distintas técnicas estadísticas de regresión o de análisis multivariante, facilitando el estudio del ciclo de vida de clientes.

Mediante la utilización de redes neuronales es posible llegar a predecir las posibles respuestas frente a determinados eventos, como las campañas comerciales y de comunicación de la empresa: ¿qué clientes son susceptibles de responder a una campaña de promoción y por qué?, ¿cuáles se verían más afectados por un cambio en la política de precios?, etc.

También se podría llegar a predecir la evolución temporal de los distintos segmentos de la cartera de clientes, facilitando la identificación de las bajas más probables para reducir de este modo el nivel de *churn* (término utilizado en el marketing para referirse a las bajas de los clientes). De este modo, la empresa podría anticiparse a las posibles bajas de clientes mediante la adopción de determinadas acciones y programas de fidelización, mejorando su índice de retención de clientes.

## **5.6 Explotación de la información registrada en Internet**

A partir del análisis de los registros de conexión a un sitio web es posible determinar cuáles son las páginas y contenidos de dicho sitio que tienen más éxito, evaluar los resultados de las campañas de promoción *on-line* (respondiendo a la pregunta de si es rentable la inversión en publicidad en determinados sitios dentro de Internet), estudiar

---

<sup>1</sup> Una técnica de venta que trata de, cuando un cliente busca comprar un producto, ofrecerle un producto del tipo que busca pero de una gama alta.

fallos y deficiencias en las conexiones (porcentaje de sesiones perdidas, transacciones incompletas...), etc., obteniendo una valiosa información que permitirá mejorar los contenidos y servicios incluidos en el sitio web.

De este modo, es posible realizar un seguimiento de las conexiones al sitio web para analizar las pautas de comportamiento general de los usuarios: ¿cuánto tiempo se dedican a consultar nuestras páginas web?, ¿se registran y nos facilitan sus datos de contacto?, ¿repiten sus visitas con frecuencia?, ¿cuáles son los momentos del día o los días de la semana con mayor actividad?, ¿desde qué países o zonas geográficas se obtienen un mayor número de visitas?, ¿responden a una determinada promoción?, ¿compran *on-line*?, etc.

Así mismo, la identificación de los visitantes de un sitio web permite crear una valiosa base de datos de marketing, que puede ser utilizada para ofrecer productos y servicios a medida, personalizar el proceso de comunicación y desarrollar otras técnicas de marketing.

También es posible definir el perfil de cada uno de los visitantes a partir del seguimiento de sus pasos dentro de un sitio web en sus distintas conexiones: qué páginas visita, cuánto tiempo le dedica a cada una de las secciones, qué productos y servicios busca dentro del catálogo de la empresa, etc. Con toda esta información los responsables del sitio podrán adaptar el contenido y la información en función de cada uno de los perfiles generados.

### 5.6.1 Análisis de los registros de actividad

Mediante el análisis de los *logs* (registros de actividad) del servidor web es posible determinar, de forma no intrusiva (es decir, sin tener que preguntárselo directamente ni incomodar al usuario), la siguiente información:

- Fecha y hora en la que tiene lugar cada visita: permite obtener la distribución semanal y distribución horaria del tráfico de visitas al sitio web.
- Dirección IP de la máquina del visitante: permite averiguar el país, el dominio y la organización a la que pertenece.
- Dirección URL de la que procede: permite detectar si el usuario proviene de una página web en la que la empresa ha insertado un *banner* u otro elemento publicitario, de una consulta por palabra clave en un buscador, de la página web de un miembro de un programa de afiliación, etc.
- Idioma, tipo de navegador y tipo de sistema operativo utilizado en el equipo del usuario que visita el sitio web.
- Páginas web solicitadas y tiempo dedicado a cada visita: facilita el análisis de qué contenidos son los más consultados y cuáles los menos interesantes dentro del sitio web, así como cuáles son los recorridos típicos por el sitio, los puntos de entrada y los puntos de salida hacia otras direcciones de Internet.
- Códigos de respuesta, estado de la conexión y cantidad de información enviada por el servidor atendiendo a cada petición del cliente.

A partir de estos datos es posible obtener otra información de interés. Así, por ejemplo, el navegador Internet Explorer solicita el icono "favicon.ico" al servidor web cada vez que el usuario decide registrar una página web en su carpeta de favoritos (*bookmarks*).

Mediante el control de este tipo de peticiones, una organización podrá conocer qué visitantes de su sitio web han decidido registrar una o varias páginas en su carpeta de Favoritos.

También es posible realizar un seguimiento de la lectura y utilización de los mensajes de correo electrónico en formato HTML, ya que el servidor web puede registrar cada una de las peticiones realizadas desde el equipo del usuario, cuando éste decide activar un determinado enlace incluido en el cuerpo del mensaje de correo electrónico. De este modo se puede obtener un conocimiento mucho más detallado de los resultados de una campaña de comunicación a través del correo electrónico.

Una empresa podría generar distintas versiones de sus páginas web, con diseños alternativos e incluyendo distintos tipos de contenidos, para poder hacer un seguimiento y evaluación en tiempo real de cuáles son las páginas web que tienen una mayor aceptación entre sus usuarios, seleccionando, en consecuencia, aquéllas que permitan mejorar la experiencia del usuario y la navegabilidad dentro del sitio web de la organización.

### 5.6.2 Técnicas de *web mining*

En lo que se refiere al análisis de los datos registrados en los registros de los servidores web, existe la posibilidad de aplicar técnicas de minería de datos a este tipo de tareas. De este modo, podemos utilizar el término de minería web (*web mining*) para referirnos a la utilización de las herramientas y técnicas de minería de datos para descubrir y extraer información sobre la utilización de los servicios de un sitio web.

Dentro de la minería web podemos distinguir las siguientes áreas de trabajo:

- **Minería de la estructura del sitio web (*website structure mining*)**, que trata de analizar cuál es la estructura real de un sitio web a través del estudio de los enlaces estáticos y dinámicos entre sus distintas páginas web, contenidos y servicios. **Resumen de sitio web**
- **Minería del contenido del sitio web (*web content mining*)**, centrada en la recopilación de datos e identificación de patrones relativos al acceso a los contenidos del sitio web por parte de sus usuarios (minería de páginas web), así como a las búsquedas que éstos realizan en Internet (minería de resultados de búsqueda).
- **Minería de la utilización del sitio web (*web usage mining*)**, que parte de los registros de los servidores web para analizar la navegación y las distintas transacciones realizadas por los usuarios, a fin de determinar sus patrones de uso: páginas más visitadas, recorridos habituales, lugar por donde comienza y por donde finaliza la visita, tiempo medio de visita, otros sitios remitentes, etc.

Por lo tanto, mediante el análisis de los datos específicos del usuario (sobre todo cuando éste se ha registrado, indicando algunos datos de tipo sociodemográfico), de los datos sobre su entorno tecnológico (navegador, sistema operativo, resolución de la pantalla, software instalado, tipo de conexión, etc.) y de los datos acerca de la utilización y navegación dentro del sitio web, es posible implantar servicios y contenidos personalizados, como los sistemas para la recomendación de los productos que mejor se ajusten a las necesidades de cada usuario.

No obstante, pese al espectacular crecimiento del número de sitios web disponibles en Internet, todavía un porcentaje muy bajo de éstos recurren a la minería web para

analizar su estructura, su contenido y su utilización, con el objetivo de poder mejorar el servicio ofrecido y la experiencia de sus usuarios.

### 5.7 Aplicaciones de apoyo a la toma de decisiones

Las herramientas presentadas en las secciones anteriores constituyen la base para el desarrollo de herramientas de apoyo a la dirección, en las que se incluyen los modelos de gestión que servirán como base para la toma de decisiones. Son variadas las aplicaciones que podemos encontrar en este ámbito:

- Modelos de consolidación empresarial.
- Planificación y control presupuestario.
- Cuadros de mando analíticos/funcionales.
- Cuadro de mando integral.
- *Reporting* de negocio o financiero.
- Análisis de correlaciones, técnicas estadísticas avanzadas o minería de datos.

Se trata de aplicaciones típicas en la mayor parte de las organizaciones, tanto públicas como privadas, que toman los datos de sus sistemas operacionales. Por esta razón, en muchas ocasiones son los propios fabricantes de los sistemas operacionales los que proponen estos modelos de gestión basándose en herramientas de *business intelligence* propias o de terceras empresas.

Con las herramientas de *business intelligence* los directivos pueden disponer de la información necesaria en muy poco tiempo y con el mínimo esfuerzo. De este modo, los directivos pueden dedicarse más al análisis de la información obtenida y no tanto a su búsqueda.

Por otra parte, dado que la información se encuentra disponible en series temporales, a partir de la acumulación de los datos del negocio y de la experiencia de la organización, es posible detectar tendencias y realizar previsiones de cara al futuro.

De este modo, podemos concluir afirmando que las herramientas de *business intelligence* facilitan la toma de decisiones estratégicas en la empresa:

- Planificación y control de la estrategia comercial y de marketing; productos, canales, política de precios, etc.
- Gestión logística y productiva.
- Planificación y control de los recursos empresariales, tanto de los recursos humanos, como de los recursos materiales y de los recursos económico-financieros.

#### 5.7.1 Cuadro de mando integral (*balanced scorecard*)

Entre las aplicaciones orientadas a la dirección, destaca en el momento actual el cuadro de mando integral o *balanced scorecard*, denominación con la que también se conoce a este tipo de aplicación. En el año 2006 un 70% de las grandes empresas declararon contar ya con esta herramienta o estar en el proceso de su implantación.



El cuadro de mando integral es un instrumento de planificación y control que favorece formular e implantar la estrategia y comprobar que se ejecuta con éxito. Sus objetivos son los siguientes:

- Concretar la estrategia en términos operativos y fijar objetivos con indicadores.
- Liderar el cambio mediante directrices de actuación y el aprendizaje por medio de la gestión de la experiencia adquirida durante las fases de diseño, implantación y control.
- Determinar las hipótesis de la estrategia mediante las relaciones causales entre acciones y objetivos.
- Guiar el esfuerzo de la organización por medio del desglose de los objetivos globales en otros específicos de cada área de gestión, promoviendo un consenso que estimule la iniciativa del personal del equipo.

Básicamente la idea del cuadro de mando integral se basa en que no se puede administrar lo que no se puede medir, por lo que es necesario establecer indicadores cuantitativos relacionados con los distintos elementos clave de la estrategia empresarial.

Al estar vinculados a la dirección general, estos indicadores suelen tener un gran nivel de agregación pues se suelen referir a la organización en su conjunto por lo que para su implementación suele recurrirse al uso de herramientas de *business intelligence*.

La forma de presentación de los indicadores suele ser muy variada e incorpora el uso de gráficas que simplifiquen la comprensión de la información, o incluso el uso de códigos semafóricos para destacar aquellos indicadores que han alcanzado los objetivos previstos de los que no.