

LAS TECNOLOGÍAS PARA LA BÚSQUEDA DE CONOCIMIENTO

Ing María del Rosario Bruera

Lic. Néstor M. Martínez

Keywords:

Data Mining, Knowledge Discovery, Data Warehouse, Business Intelligence, Marketing Directo.

La era de la información.

En los últimos años ha crecido exponencialmente la necesidad de procesar rápidamente la información de negocios para acelerar la toma de decisiones. Sin embargo, herramientas de uso habitual como las hojas de cálculo o eventualmente los lenguajes de consulta a las bases de datos (SQL) no resultan suficientemente eficaces para encontrar rápidamente las respuestas a preguntas de negocio tales como:

¿Quiénes son mis mejores clientes?

¿Cuánto debo invertir en retener a un cliente?

¿Qué productos se venden en forma conjunta?

¿Cuáles de mis clientes están a punto de abandonar mi negocio?

¿Quiénes responderán a una campaña a través de un mailing?, etc, etc

Por otra parte, Internet, el comercio electrónico, la banca electrónica y la posibilidad de registrar y mantener en línea las transacciones de los clientes proveen una fuente de datos de inmenso valor para el desarrollo competitivo de los negocios.

Las nuevas tecnologías de análisis de datos – o búsqueda de conocimiento - resultan entonces una ventaja evidente para aquellos que las adoptan e integran a sus procesos comerciales.

Con información apropiada se disminuyen costos, se mejora la relación con el cliente, se orienta el diseño de los productos, se optimizan los recursos de las empresas.

Los “almacenes de datos” o “datawarehouses”.

Desde los años 90, las organizaciones comenzaron a percibir que el uso **estratégico** de los datos para orientar decisiones de negocios es una disciplina totalmente diferente de los usos **operativos** de estos mismos datos destinados – a través de los sistemas informáticos – a soportar el funcionamiento y la administración de los procesos comerciales.

Estas necesidades derivaron en la construcción de almacenamientos de datos corporativos o **datawarehouses** que integran los datos operativos de los sistemas transaccionales en almacenes de fácil acceso preparados para un procesamiento analítico.

Un datawarehouse es :

“Una copia de los datos transaccionales optimizada para consultas complejas y análisis”

Muchas organizaciones han dedicado esfuerzo y destinado recursos en los últimos años a la construcción de sus datawarehouses y a la recolección de sus datos. Sin embargo, muy pocas han logrado demostrar si la inversión realizada en tecnología tuvo el retorno que la justificara. Esto sucede en parte porque el verdadero valor del datawarehouse solamente se percibe a través de la **explotación** de su contenido.

Las herramientas para búsqueda de conocimiento.

Esta explotación del contenido del datawarehouse se realiza a través de herramientas que se agrupan en cuatro categorías básicas:

- . Herramientas para generación de informes y consultas
- . Herramientas de visualización
- . Herramientas OLAP
- . Herramientas de minería de datos

Las herramientas de generación de informes y consultas asisten al usuario en la confección de resúmenes de información de gestión que pueden luego integrarse a un tablero de control.

Las herramientas de visualización **mapean** la información multidimensional contenida en el datawarehouse en el plano de la pantalla. Las más conocidas son las de información geográfica (o GIS) que permiten, por ejemplo, señalar con colores diferentes las áreas que rodean a las sucursales mostrando, por ejemplo, la densidad de clientes alrededor de ellas e identificando potenciales lugares para la ubicación de nuevas sucursales.

Las herramientas **OLAP** (On Line Analytical Processing) construyen vistas multidimensionales de los datos del datawarehouse y las almacenan en **cubos**. Estos cubos son representaciones tabulares de los datos que incluyen cálculos y sumalizaciones. Las herramientas OLAP automáticamente permiten recalcular los cubos según diferentes niveles de agregación jerárquicos (drill-down). Por ejemplo, un resumen por región geográfica puede luego abrirse por provincia, localidad, sucursal, etc. llegando hasta el mínimo nivel de análisis soportado por el datawarehouse.

Las herramientas OLAP son sin duda, las grandes aliadas para la explotación de los datos, pero aún se puede llegar más lejos incorporando una nueva generación de herramientas que se conoce con el nombre de **Minería de Datos**.

A diferencia de las herramientas OLAP, la minería de datos puede proporcionar modelos no solamente **descriptivos** , sino también **predictivos** de los patrones ocultos en los datos. Asisten al usuario para el análisis de grandes volúmenes de información o de patrones complejos de comportamiento que involucran una gran cantidad de variables. Desde un punto de vista metodológico, se apoyan en algoritmos derivados de la estadística y de la inteligencia artificial convenientemente adaptados para objetivos comerciales.

Sin embargo, la aplicación de estas nuevas herramientas aún **no ha podido automatizar:**

- La elección de los problemas de negocio candidatos para su aplicación
- La identificación y recolección de los datos que contienen la información buscada
- El “masajeo” y tratamiento de los datos que posibilita la búsqueda de patrones
- El plan de acciones que apoyándose en los resultados del modelo produzca el retorno de la inversión
- La medición del éxito de las acciones realizadas a partir de los resultados

Es por esto que los proyectos de búsqueda de conocimiento orientada a la gestión comercial requieren para su éxito de un importante soporte de capacitación de los usuarios de las técnicas así como también de los responsables de la implantación de la tecnología.

Aplicaciones habituales.

Algunas de las aplicaciones habituales de estas nuevas tecnologías se incluyen en la lista siguiente, que naturalmente, no exhaustiva y mucho menos completa:

Retención de clientes: Modelos que permiten asignar a cada cliente una probabilidad de abandono según varíen sus patrones de consumo.

Ventas y servicio al cliente: Los *modelos de canasta de productos* pueden emplearse para ofrecer la mejor oferta, al cliente adecuado, en el momento oportuno. Incrementan el cross-selling, permiten la selección de ofertas de catálogo o de páginas Web, sirven de base para la ubicación de productos en el local, etc, etc.

Marketing : Marketing depende casi totalmente de los análisis de la información de negocio para ejecutar campañas, optimizar tasas de respuesta, asignar valor a los clientes y determinar perfiles de oferta. Los costos de marketing directo son inversamente proporcionales a la **bondad** de los modelos construidos. Cuanto más se ajusta el modelo, menos costosas resultan las campañas.

Análisis de riesgo y detección de fraudes : aplicaciones propias de los sistemas financieros, de tarjetas de crédito y últimamente de compañías de telefonía y comunicaciones permiten detectar por ejemplo transacciones anómalas según el patrón habitual del cliente y anticiparse al reclamo.

Mega-proyectos o proyectos posibles?

La tarea de búsqueda de conocimiento solamente tiene sentido si el valor de los resultados supera a la inversión realizada para obtenerlos. En el contexto de las pequeñas y medianas empresas muchas veces se cree que para explotar estratégicamente los datos de la compañía resulta necesaria una gran inversión que incluya el desarrollo de un mega proyecto de datawarehouse previo antes de poder obtener algún beneficio. Nada más alejado de la realidad que esta conclusión a-priori. Disponer de un datawarehouse sin duda facilita las tareas

de análisis pero de ninguna manera es una condición excluyente. Los mejores proyectos de explotación de datos se consiguen a partir de necesidades puntuales y acotadas con resultados medibles en el corto plazo. Existen en el mercado soluciones de hardware-software-metodología de costos razonables y consistentes con las necesidades de una PyME por lo que cualquier empresa sin importar su tamaño puede beneficiarse con estas tecnologías en la medida que integre sus objetivos comerciales con las nuevas herramientas de búsqueda de conocimiento.

Los autores.

María del Rosario Bruera es Ingeniera, egresada de la Universidad Nacional de Rosario. Desde 1978 trabaja en aplicaciones informáticas, desarrollo de software y modelización de datos sobre distintas plataformas. Docente Universitaria en Ingeniería de Software, Metodología de Programación, Estadística y Bases de Datos. Ex integrante del Conicet como profesional de apoyo a la investigación en informática y estadística. Coordinadora de proyectos de capacitación para empresas y representantes de productos de software. Es miembro del CRISP-DM Special Interest Group, organización internacional dedicada a la standarización de los procesos correspondientes al Data Mining Life Cycle.

Néstor Martínez es Estadístico, egresado de la Universidad Nacional de Rosario. Especialista en Análisis de Datos y Data Mining aplicados a la Economía y a las Ciencias Sociales con importantes conocimientos Tecnología Informática. Tiene amplia experiencia en diferentes entornos de trabajo: investigación, académicos, organismos públicos y compañías privadas relacionados con diferentes industrias: farmacéuticas, automóviles, bancos, seguros, tarjetas de crédito, telecomunicaciones, consumo masivo, AFJP, etc. Ex consultor en Estadística e Informática del Programa de las Naciones Unidas para el Desarrollo. Ex Statistics & Quality Control Manager de IMS Argentina, Uruguay, Paraguay y Bolivia.