

**UNIVERSIDAD NACIONAL DE INGENIERÍA**  
**FACULTAD DE INGENIERÍA QUÍMICA Y MANUFACTURERA**



**“OPTIMIZACIÓN Y CONTROL DE PROCESOS UTILIZANDO EL  
CONCEPTO DE INTELIGENCIA DE NEGOCIOS”**

**INFORME DE SUFICIENCIA**

**PARA OPTAR EL TÍTULO PROFESIONAL DE:  
INGENIERO QUÍMICO**

**POR LA MODALIDAD DE ACTUALIZACIÓN DE  
CONOCIMIENTOS**

**PRESENTADO POR:**

**JAVIER GUILLERMO GONZALES SANDOVAL**

**LIMA – PERÚ**

**2004**

## RESUMEN

Hoy en día la capacidad de acceso a la información crece a pasos agigantados pero no es fácil asimilarla. Muchas veces las organizaciones no saben obtener información valiosa de toda la cantidad ingente de datos que tienen, pero saben que actualmente el conocimiento constituye el activo más valioso de cualquier organización. El éxito de las organizaciones, así como su supervivencia, depende básicamente de que este conocimiento pueda preservarse y utilizarse eficientemente.

La Inteligencia de Negocios facilita la información en el momento preciso a cada uno de las partes que constituyen la empresa, para que así se obtenga un máximo rendimiento en sus actividades. Esto se debe a que permite tomar la gran cantidad de datos que genera diariamente la empresa, y transformarlos en la información que necesita para monitorear su salud financiera y operativa, y para identificar rápidamente tendencias y tomar decisiones en forma ágil y proactiva para superar a los competidores.

Entre las principales herramientas de la Inteligencia de Negocios podemos mencionar a la Prospectiva Tecnológica y la Minería de Datos. El presente estudio se centra en la aplicación de la Minería de Datos en Ingeniería Química.

La Minería de Datos, también conocida como descubrimiento del conocimiento (Knowledge Discovery), es el proceso asistido por computadora que busca, extrae y analiza enormes cantidades de datos para descubrir patrones y derivar su significado. Estos patrones y estructuras de información pueden resultar muy valiosos para la toma de decisiones en la industria, los bancos, las empresas comerciales y empresas en general.

En la sección 1, "Introducción", se realiza una aproximación a la Inteligencia de Negocios y su aplicación exitosa en la predicción del comportamiento humano y de procesos administrativos.

En la sección 2, "Desarrollo de conceptos y técnicas", se profundizan los conceptos de Inteligencia de Negocios y Minería de datos. También se analiza la planta como un gran almacén de información aun no explotado, y se presentan aplicaciones directas de la Minería de Datos en los procesos manufactureros y sus beneficios.

En la sección 3, "Desarrollo del tema", se presenta la elaboración de un modelo de predicción y estimación empleando Minería de Datos. Luego se compara el modelo con modelos desarrollados empleando Redes Neuronales, para demostrar sus ventajas:

1. No requiere una relación funcional entre datos de entrada y de salida
2. Se basa en el análisis de información previa existente
3. El grado de confiabilidad del modelo determina directamente su validez

También se aplica el modelo en procesos específicos de la industria, como creación de sensores virtuales para el control de procesos, determinación de fallas en equipos, y determinación de parámetros óptimos de operación para elevar calidad o reducir costos.

Finalmente, en la sección 4 "Conclusiones y Recomendaciones", presentamos los beneficios de la aplicación de estas herramientas en los procesos manufactureros, y como su uso permite a la información de manufactura agregar valor a la empresa. También se incluyen recomendaciones para la implementación de proyectos de este tipo.

# ÍNDICE

<b>1</b>	<b>INTRODUCCION.....</b>	<b>7</b>
<b>2</b>	<b>DESARROLLO DE CONCEPTOS Y TÉCNICAS .....</b>	<b>9</b>
2.1	CONCEPTOS DE INTELIGENCIA DE NEGOCIOS.....	9
2.1.1	<i>¿Qué es Inteligencia de Negocios? .....</i>	<i>11</i>
2.1.2	<i>¿A quien va dirigida?.....</i>	<i>13</i>
2.1.3	<i>¿Qué nos permite? .....</i>	<i>15</i>
2.1.4	<i>Aplicaciones .....</i>	<i>15</i>
2.1.5	<i>Beneficios de su aplicación: Mayor Rentabilidad .....</i>	<i>19</i>
2.1.6	<i>Data Warehousing y Data Marts .....</i>	<i>20</i>
2.2	MINERÍA DE DATOS: HERRAMIENTA DE LA INTELIGENCIA DE NEGOCIOS PARA DESARROLLO DE MODELOS DE ESTIMACIÓN Y PREDICCIÓN .....	23
2.2.1	<i>¿Qué es Minería de Datos? .....</i>	<i>24</i>
2.2.2	<i>¿Cuándo aplicamos Minería de Datos? .....</i>	<i>25</i>
2.2.3	<i>Minería de Datos frente a OLAP y DSS .....</i>	<i>29</i>
2.2.4	<i>Usos frecuentes de la Minería de Datos .....</i>	<i>33</i>
2.2.5	<i>Fases de un proyecto de Minería de Datos.....</i>	<i>39</i>
2.2.6	<i>Aprendiendo de los datos históricos .....</i>	<i>41</i>
2.2.7	<i>Prediciendo el futuro .....</i>	<i>44</i>
2.3	TRANSFORMANDO LA PLANTA EN UN ALMACÉN DE INFORMACIÓN.....	45
2.3.1	<i>El valor de la información de planta para la empresa .....</i>	<i>46</i>
2.3.2	<i>¿Cómo agrega valor a la empresa la información de manufactura?.....</i>	<i>47</i>
2.3.3	<i>La integración entre la planta y la administración de la empresa: Los sistemas MES (Manufacturing Execution Systems) .....</i>	<i>49</i>
2.4	APLICACIÓN DE MINERÍA DE DATOS EN LOS PROCESOS DE PLANTA .....	50
2.4.1	<i>Administración de flujo de órdenes .....</i>	<i>51</i>
2.4.2	<i>Administración de flujo de producción y materiales .....</i>	<i>52</i>
2.4.3	<i>Administración de los recursos de la empresa .....</i>	<i>53</i>
2.4.4	<i>Sincronización de la cadena de valor .....</i>	<i>54</i>

2.4.5	<i>Otras aplicaciones de la Minería de Datos</i> .....	56
<b>3</b>	<b>DESARROLLO DEL TEMA</b> .....	<b>57</b>
3.1	MODELAMIENTO TOPOLÓGICO BASADO EN CASOS (TCBM).....	57
3.1.1	<i>El problema típico a resolver</i> .....	58
3.1.2	<i>La solución: El modelo TCBM</i> .....	59
3.1.3	<i>Características del TCBM</i> .....	60
3.1.4	<i>Proceso de construcción del modelo</i> .....	62
3.2	CASO PRÁCTICO: DESARROLLO DE UN MODELO PARA DETERMINAR LA CONVERSIÓN EN LA PRODUCCIÓN DE ÁCIDO SULFÓNICO .....	63
3.2.1	<i>Paso 1: Cargar los datos al software</i> .....	65
3.2.2	<i>Paso 2: Separación de los datos</i> .....	66
3.2.3	<i>Paso 3: Modelamiento</i> .....	68
3.2.4	<i>Paso 4: Verificación</i> .....	72
3.2.5	<i>Paso 5: Estimación</i> .....	73
3.3	COMPARACIÓN ENTRE METODOLOGÍAS DE MODELAMIENTO DE CAJA NEGRA.....	76
3.3.1	<i>Modelamiento empleando Redes Neuronales Artificiales</i> .....	76
3.3.2	<i>Modelamiento topológico basado en casos</i> .....	79
3.3.3	<i>El modelamiento topológico basado en casos resuelve los problemas del modelamiento basado en redes neuronales</i> .....	81
3.4	APLICACIÓN PRÁCTICA DE MODELOS TCBM EN LA AUTOMATIZACIÓN Y CONTROL DE PROCESOS.....	82
3.4.1	<i>Sensor virtual para la medición de concentración de NOx</i> .....	82
3.4.2	<i>Determinación de causa de falla para mantenimiento de equipos (empresa Weyerhaeuser)</i> .....	85
3.4.3	<i>Determinación de parámetros óptimos de operación para mejorar la calidad de productos</i> .....	87
<b>4</b>	<b>CONCLUSIONES Y RECOMENDACIONES</b> .....	<b>89</b>
4.1	CONCLUSIONES.....	89
4.2	RECOMENDACIONES .....	91

<b>5</b>	<b>BIBLIOGRAFIA.....</b>	<b>93</b>
<b>6</b>	<b>APÉNDICE: CASOS DOCUMENTADOS DE APLICACIÓN DE MINERÍA DE DATOS A PROCESOS INDUSTRIALES.....</b>	<b>94</b>

# 1 Introducción

Cada vez más, y gracias al desarrollo de las comunicaciones y la implementación de redes informáticas, la cantidad de información que fluye en las empresas ha aumentado considerablemente. Sin embargo, el tiempo para asimilarla es menor. Para que las empresas trabajen eficientemente se requiere el mínimo de información necesaria y tiene que ser presentada en forma fácil, de esta manera se podrá interpretar y manejar con mayor eficiencia. Si el objetivo es que la empresa sea exitosa, entonces la empresa requiere agilidad, es decir, la capacidad de anticipar las tendencias emergentes y las oportunidades, a fin de transformar de manera instantánea la estrategia y los procesos internos para reenfocar sus recursos y aprovechar al máximo las nuevas posibilidades.

La Inteligencia de Negocios es un conjunto de herramientas que proveen a la empresa la agilidad que requiere para enfrentar con éxito su presente y futuro. Entre las principales herramientas de la Inteligencia de Negocios podemos mencionar a la Prospectiva Tecnológica y la Minería de Datos.

La Prospectiva Tecnológica es un proceso colectivo de análisis y comunicación para identificar los componentes probables de escenarios futuros: las proyecciones tecnológicas, sus efectos sociales y económicos, los obstáculos y las fuerzas que operan a favor. La Minería de Datos, también conocida como descubrimiento del conocimiento (Knowledge Discovery), es el proceso asistido por computadora que busca, extrae y analiza enormes cantidades de datos para descubrir patrones y derivar su significado. La Minería de Datos involucra tanto la descripción de lo sucedido (pasado) como la predicción de tendencias futuras.

La Prospectiva realiza un análisis externo (exógeno) de la organización, permite construir una visión a largo plazo, y orientar a la organización hacia ese futuro. La Minería de Datos realiza un análisis interno (endógeno) de la organización, permite saber que está sucediendo y predecir lo que sucederá basado en los datos propios pasados.

Con el avance de las tecnologías de información, los sistemas de almacenamiento de datos, la industria y empresas han dedicado gran parte de sus esfuerzos a construir grandes bases de datos con el solo objetivo de llevar un registro histórico de los hechos ocurridos. Sin embargo, en muchos casos, toda esta enorme cantidad de datos no es aprovechada adecuadamente, sabiendo además que en ella se puede perder información importante (patrones de comportamiento, modos de operación, información útil para descubrir fallos, tendencias, etc.), la cual puede ser "rescatada" a través de metodologías basadas en Minería de Datos.

La Minería de Datos permite tomar estos grandes volúmenes de datos generados diariamente y transformarlos en la información que necesita la organización para monitorear su salud financiera y operacional, y para identificar rápidamente tendencias y tomar decisiones en forma ágil y proactiva para superar a los competidores.

El campo de aplicación de la Minería de Datos es amplio: control, optimización y supervisión de procesos industriales, control de calidad, modelado e identificación de sistemas, obtención de tendencias de la bolsa de valores, correlación entre indicadores financieros, diagnóstico de enfermedades, determinación de los efectos de un medicamento, clasificación de señales biomédicas, predicción de ventas, planificación de campañas publicitarias, detección de fraudes, detección de evasión de impuestos, detección de redes de narcotráfico, hallazgos de patrones de comportamiento, etc.

Para obtener buenos resultados es necesario comprender que la Minería de Datos no se basa en una metodología estándar y genérica que resuelve todo tipo de problemas, sino que consiste en una metodología dinámica e iterativa que va a depender del problema planteado, de la disponibilidad de datos, del conocimiento de las herramientas, de la metodología desarrollada y de los requerimientos y recursos de la empresa.

El presente trabajo se centra en la aplicación de la Minería de datos en la optimización y control de procesos manufactureros.

## **2 Desarrollo de Conceptos y Técnicas**

### ***2.1 Conceptos de Inteligencia de Negocios***

Nos gusta pensar que la empresa es un esquema racional en el que los responsables de la toma de decisiones primero recopilan toda la información disponible, luego hacen un análisis completo de las condiciones del negocio y evalúan los riesgos y beneficios potenciales de sus conclusiones, antes de elegir una ruta de acción.

Los estudios demuestran que, con frecuencia, este no es el caso. De acuerdo con un estudio reciente realizado en los Estados Unidos y el Reino Unido, cerca de 9 de cada 10 gerentes indicaron que la intuición es más importante que los hechos concretos en el 75% de las decisiones que toman. Adicionalmente, más del 60% de los ejecutivos indicaron que, por lo general, no cuentan con suficiente información cuando toman decisiones. El hecho es que, a menudo, la toma de decisiones en las empresas es prácticamente un proceso subjetivo, guiado más por presentimientos que por un entendimiento claro del panorama de negocios.

Esto puede ser contra productivo, debido al crecimiento explosivo en el volumen de información que se recopila y almacena en las empresas hoy en día. Esto significa que en lugar de aumentar nuestro entendimiento del panorama de negocios, con frecuencia toda la información sirve para agregar incertidumbre. El problema es que incluso para las empresas técnicamente avanzadas, la información se recolecta estrictamente por unidades de negocios en sistemas aislados de almacenamiento, incluyendo bases de datos, servidores de archivos y discos duros de las computadoras de las personas que toman las decisiones.

Esto crea una combinación compleja. Por un lado, los responsables de la toma de decisiones a escala empresarial son inundados con grandes cantidades de datos,

pero se les obliga a tomar decisiones sin tener acceso a información crítica que pueda estar atrapada en sistemas a los cuales no tienen acceso. Como resultado, las decisiones se toman con una perspectiva limitada y a menudo imprecisa, de las condiciones empresariales. Esto ha dado lugar a la ironía fundamental de la revolución de la información: procesamos tanta información de negocios, pero tenemos muy poca Inteligencia de Negocios. Incluso las empresas que han implementado costosas tecnologías de *Data Warehouse*<sup>1</sup> para administrar mejor los datos que recolectan, carecen de las herramientas integradas y de la infraestructura necesaria a escala empresarial para extraer de toda esa información un beneficio tangible para su negocio.

El problema se complica aún más por el hecho de que el ambiente de negocios actual exige que se tomen decisiones rápidas. Con la llegada de Internet y del comercio electrónico, se han acelerado todos los aspectos de los negocios. Hoy es primordial tener la capacidad de responder rápidamente a los cambios del mercado. La tarea de filtrar y procesar datos y entregar informes a ejecutivos de la alta gerencia era antes dominio de especialistas de investigación. Pero esto ha cambiado, ahora cada empleado en todos los niveles de una empresa tiene la responsabilidad de recopilar y analizar la información.

Por lo tanto, el reto es proporcionar a cada funcionario de su empresa una vista enfocada de todos los datos disponibles, junto con las herramientas que necesitan para sintetizar los datos y transformarlos en información que permita tomar decisiones de negocios acertadas y oportunas.

---

<sup>1</sup> El término *Data Warehouse* es de difícil traducción. No existe un término único aceptado comúnmente, por lo que en el presente trabajo se prefiere mantenerlo en su idioma original. Se define habitualmente como sistema que permite el almacenamiento en un único entorno de la información histórica e integrada de la empresa, y que refleja los indicadores claves asociados a los negocios de la misma. En la sección 2.1.6 *DataWarehousing* y *Data Marts* ampliamos su definición.

La respuesta a estas necesidades de información de las empresas de hoy es la **Inteligencia de Negocios**. Los sistemas de Inteligencia de Negocios integran datos de todas las unidades de la empresa y se los entregan a las diferentes personas de la organización en una forma estructurada y lista para su análisis. La Inteligencia de Negocios abarca más allá del análisis, integrando herramientas que dan a los profesionales la capacidad de convertir sus percepciones en decisiones, transformando el entendimiento en medidas concretas en tiempo real.

Cuando la Inteligencia de Negocios se aborda estratégicamente, ofrece la oportunidad de eliminar las conjeturas que todavía forman parte del proceso de toma de decisiones empresariales. Las soluciones obtenidas por la aplicación de Inteligencia de Negocios pueden reducir costos, aumentar la flexibilidad, mejorar la satisfacción a clientes, optimizar procesos, proporcionar mayores utilidades y dar una ventaja competitiva distintiva en el mercado global.

### **2.1.1 ¿Qué es Inteligencia de Negocios?**

En los términos más simples, la inteligencia de Negocios es un esfuerzo para captar y analizar los “datos del negocio” para comprender la tendencia de los mercados, la empresa y clientes con mayor claridad, a fin de mejorar los procesos de negocios y competir con más efectividad. Se debe entender como “datos del negocio” al conjunto de indicadores clave que permiten medir y controlar el desempeño de la empresa.

El concepto se originó hace más de 30 años, en ese entonces las aplicaciones de inteligencia de Negocios utilizaban herramientas de consulta e informe de los grandes computadores (mainframes) para realizar análisis de productos y mercados. La Inteligencia de Negocios en esa época requería de implementaciones costosas y de un conocimiento de herramientas muy complejas y sistemas cerrados de bases de datos relacionales, que de ninguna manera eran fáciles de usar. Además, demoraba mucho tiempo generar los informes. Para

ejecutar las consultas, los sistemas tenían que salir de operación, por lo que la mayoría de los informes se ejecutaban en lotes. Eso significaba que con frecuencia se retrasaban los análisis y los informes se basaban en datos desactualizados. Así mismo, las limitaciones en el almacenamiento obligaban a los gerentes de bases de datos a borrar frecuentemente los datos de los sistemas, eliminando la posibilidad de ejecutar análisis que mostraran la forma en que evolucionaban las condiciones del negocio con el tiempo.

A principios de los 80 se introdujeron los Data Warehouses, con el fin de agilizar y simplificar el acceso a los datos y mejorar las capacidades de análisis. Los Data Warehouses permitieron a los gerentes de bases de datos extraer datos transaccionales y almacenarlos en depósitos que se podían optimizar para su análisis. Los Data Warehouses abrieron la ventana al análisis en serie y redujeron la demora entre la recopilación y el análisis de los datos.

El siguiente paso importante llegó con la tecnología de Procesamiento de Análisis en Línea (OLAP), a finales de los 80. OLAP permite a las empresas tener acceso a datos complejos casi en tiempo real, con un mejor rendimiento en las consultas y una construcción de consultas más simple que las que proporcionan las bases de datos relacionales tradicionales.

No, obstante, durante la mayor parte de los 90, la inteligencia de Negocios a escala empresarial continuó estando fuera del alcance de casi todas las empresas, a excepción de las grandes corporaciones. Las barreras incluían el costo abrumador de adquirir el hardware y las aplicaciones necesarias para crear y ejecutar sistemas de Data Warehousing y análisis, el extenso tiempo de espera necesario para ensamblar y construir aplicaciones de análisis, el constante problema de integrar datos de sistemas existentes y cerrados, así como el alto precio del mantenimiento y la administración continua de los sistemas de inteligencia de Negocios.

Los grandes cambios tecnológicos de los últimos tres años han colocado las soluciones de inteligencia de Negocios al alcance de una mayor cantidad de empresas. El costo del hardware necesario para ejecutar sistemas de Inteligencia de Negocios se ha reducido sustancialmente, a la vez que la capacidad de almacenamiento y el poder de procesamiento, así como la velocidad de estos sistemas, han mejorado drásticamente. Tecnologías como OLAP, que en algún momento fueron aplicaciones independientes y costosas, ahora pueden adquirirse por precios muy bajos o incluso sin costo alguno. Por otro lado, la definición de nuevos estándares de Internet han reducido en gran medida las dificultades de integrar datos de fuentes que antes eran incompatibles.

Estos cambios significan para las grandes empresas que las capacidades analíticas de las aplicaciones de Inteligencia de Negocios puedan aplicarse a escala global. Al mismo tiempo, también permiten que las empresas de todo tipo y tamaño, sobre todo las PYMES, tengan a su alcance estas herramientas de análisis.

### 2.1.2 ¿A quien va dirigida?

*“Consideramos que las empresas exitosas del siglo XXI serán aquellas que capitalicen la enorme riqueza de los datos recolectados por sus sistemas de procesamiento de transacciones en línea y las aplicaciones emergentes de comercio electrónico, para poder crear lealtad entre sus clientes y operar sus negocios más eficientemente”, señala First Analyst Securities Corporation. “Creemos que las herramientas y aplicaciones de Inteligencia de Negocios son cruciales”.*

Por lo tanto, la Inteligencia de Negocios va dirigida a toda empresa que desea ser exitosa en el entorno actual. La mejor forma de reconocer si se puede o no aplicar Inteligencia de Negocios a una empresa en respondiendo a esta serie de preguntas:

¿Tiene problemas para crear un panorama claro de toda su organización?

- ¿Tiene informes de varios sistemas operacionales que no concuerdan?
- ¿Su departamento de informática está sobrecargado con solicitudes de informes para los ejecutivos?
- ¿Cuenta con unidades de negocios que han contratado a su propio personal de informática para escribir sus programas de informes?
- ¿No está seguro qué productos y clientes son los más importantes para su empresa?
- ¿Está perdiendo participación de mercado con su competencia?

Si la respuesta a cualquiera de estas preguntas es afirmativa, la empresa puede obtener beneficios instantáneos implementando sistemas de Inteligencia de Negocios más eficientes. La Inteligencia de Negocios provee a la empresa la capacidad de trazar mucho más que un plan de acción basado en comparaciones de proyecciones mensuales de ventas. Permite, por ejemplo, consolidar en tiempo real datos de los puntos de venta con información de productos, promociones y tendencias del mercado y luego proporcionar esa información instantáneamente a los encargados de tomar las decisiones que están facultados para ejecutar el trabajo.

Con la Inteligencia de Negocios, la empresa puede discernir tendencias que representan nuevas oportunidades, anticipar problemas potenciales y hacer ajustes antes de que se conviertan en un problema real. Las soluciones de Inteligencia de Negocios proporcionan a los profesionales la capacidad de profundizar en los detalles del negocio, para evaluar rápidamente tendencias, patrones y relaciones, luego utilizar esa Inteligencia para alterar el curso tomado en horas o minutos, en lugar de semanas o meses.

En la era digital, las capacidades que ofrece la Inteligencia de Negocios será la diferencia entre el éxito y el fracaso.

### 2.1.3 ¿Qué nos permite?

La Inteligencia de Negocios provee una visión holística de la empresa, mediante el monitoreo de la salud financiera y operacional de la Organización. Como se describe en la figura 2-1, permite que la empresa pueda alinear estrategias, expandir mercados, mejorar eficiencias y mantener clientes.

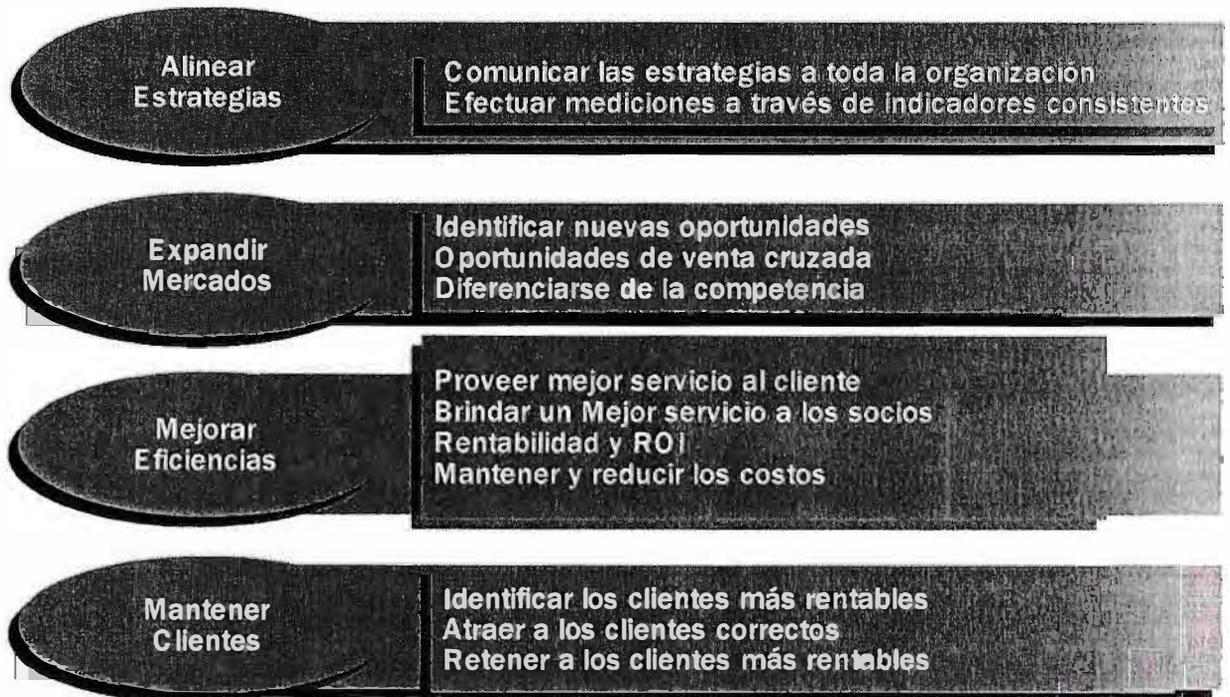


Figura 2-1 La importancia de la Inteligencia de Negocios en la empresa

### 2.1.4 Aplicaciones

La proliferación de las soluciones empresariales, tales como Planeamiento de Recursos Empresariales - ERP<sup>2</sup>, Administración de la Relación con Clientes

<sup>2</sup> ERP – Enterprise Resource Planning (Planeamiento de Recursos Empresariales), soluciones informáticas integrales para empresas. Permiten administrar e integrar las diferentes áreas (Finanzas, Recursos Humanos, Logística, Producción, etc.).

(CRM<sup>3</sup>) y la Administración de la Cadena de Suministro, entre otras, ha facilitado la recopilación de la información de negocios. En una empresa típica, cada departamento recolecta datos: datos sobre las compras de sus clientes, cuándo y cómo lo compran, con qué rapidez pagan y con qué frecuencia llaman al grupo de soporte al cliente; datos del rendimiento de la efectividad operativa, la efectividad de manufactura y el status de los inventarios; datos de mercadeo sobre su competencia y datos demográficos sobre mercados cambiantes; información de proveedores y socios; además de datos de la Web y de comercio electrónico. Dependiendo del tipo de negocio, es posible que la empresa recopile proyecciones de consumo de energía, se suscriba a datos climatológicos o lleve un control de la información de los cambios reglamentarios.

En conjunto, toda esa información podría ofrecer una perspectiva detallada y precisa de cada aspecto del negocio. Las probabilidades de que esto no suceda así son muy altas. Para la mayoría de empresas grandes, la infraestructura de información consiste en una combinación de sistemas existentes y cerrados. Los problemas de integración y compatibilidad causan que los datos tiendan a permanecer atrapados en los sistemas del departamento que los recolecta. El problema es que en la mayoría de las empresas un departamento puede no conocer los datos que otros departamentos recolectan. Por lo general, cuando los datos finalmente se resumen y están disponibles para los ejecutivos que toman las decisiones, están incompletos y desactualizados.

La Inteligencia de Negocios coloca estos datos al alcance de todos en la empresa, proporcionando los medios para extraerlos de aplicaciones operativas, convertirlos en un formato estándar y luego almacenarlos en una localidad central optimizada tanto para una entrega rápida de información resumida, como una consulta y análisis más detallados. El resultado es una vista inmediata,

---

<sup>3</sup> CRM – Customer Relationship Management (Administración de la relación con clientes), soluciones informáticas de apoyo a la interacción de la empresa con sus clientes.

personalizada y relevante de toda la información disponible, la cual va más allá de las capacidades básicas de informe de los sistemas basados en transacciones, permitiendo una toma de decisiones más rápida y mejor informada en cada nivel de la empresa.

En cualquier empresa, las decisiones forman parte de alguna de estas tres categorías: estratégica, táctica u operativa. Las decisiones estratégicas se enfocan en la dirección del negocio a largo plazo y las toman los ejecutivos de la alta gerencia. Las decisiones tácticas las realizan los gerentes de nivel medio y se enfocan en la planeación de proyectos y producción. A nivel operativo, los profesionales de las unidades de negocio toman las decisiones cotidianas que requieren para ejecutar los planes.

Cada tipo de decisión requiere una forma diferente de análisis. Los planeadores estratégicos, preocupados con los objetivos a largo plazo, analizan datos resumidos sobre el rendimiento interno y datos externos de la competencia, para explorar escenarios potenciales de negocios que les ayuden a comprender el valor de mercados nuevos o los beneficios de adquisiciones.

Los gerentes de nivel medio tienen que tener la capacidad de profundizar hasta llegar a los niveles tácticos de las ventas y producción para generar presupuestos, proyecciones de ventas y crear planes de proyectos y producción. Los profesionales individuales analizan los datos transaccionales en tiempo real para asegurar que el rendimiento vaya de acuerdo con lo planeado, hacer los ajustes necesarios para resolver problemas a corto plazo y aprovechar las nuevas oportunidades.

Los sistemas actuales de Inteligencia de Negocios proporcionan un nivel de acceso sin precedentes a los datos críticos necesarios para cada tipo de análisis, conjuntamente con un medio para integrar datos transaccionales en tiempo real a los procesos de toma de decisión, tanto a nivel estratégico como táctico. La

capacidad de incorporar datos en tiempo real a los modelos de planeación que utilizan los ejecutivos y los gerentes de nivel medio, puede mejorar drásticamente el valor de las proyecciones y los planes a mediano y largo plazo. La inteligencia de Negocios también permite a la empresa afinar la toma de decisiones cotidiana, asegurando que cada grupo operativo tenga acceso a la información necesaria para contestar preguntas específicas de las tareas. Cada departamento es responsable de un conjunto distinto de parámetros y mediciones. La Inteligencia de Negocios proporciona los medios para dar a cada grupo una vista apropiada de los datos corporativos que necesita para tener éxito. Así, las aplicaciones más frecuentes están dirigidas a las siguientes áreas de la empresa:

***Ventas y mercadeo:*** La Inteligencia de Negocios ofrece nuevas y poderosas herramientas para comprender las necesidades del cliente y responder a las nuevas oportunidades de mercado. Con sistemas integrales de Inteligencia de Negocios, los análisis de mercadeo pueden medir el impacto de los precios y promociones, permitirle dirigirse a segmentos de clientes de manera más precisa, analizar patrones de compra para aprovechar oportunidades de ventas de productos asociados y desarrollar en tiempo real una relación de mercadeo personal con cada uno de sus clientes.

***Desarrollo de productos:*** La Inteligencia de Negocios ofrece a los equipos de desarrollo de productos acceso a datos críticos de clientes y del mercado, conjuntamente con la información esencial de proveedores necesaria para realizar análisis precisos de las relaciones costo/beneficio de las características y materiales.

***Operaciones:*** Abarcando desde el control de calidad y la administración de inventarios hasta el planeamiento de la producción y la optimización de procesos productivos, la Inteligencia de Negocios proporciona un mecanismo para analizar el rendimiento de cualquier proceso operativo.

**Finanzas:** La Inteligencia de Negocios proporciona a los planificadores financieros acceso inmediato a datos en tiempo real, ofreciendo así un nuevo beneficio para las operaciones financieras, incluyendo presupuestos y proyecciones.

**Atención al cliente:** Las herramientas de Inteligencia de Negocios le permiten evaluar con precisión el valor de los segmentos de mercado y de clientes individuales, a la vez que le ayuda a retener a los clientes más rentables.

**Relaciones con proveedores:** La Inteligencia de Negocios aprovecha la integración en línea de proveedores y socios, para proporcionar nuevos niveles de análisis sobre el desempeño de los proveedores, nuevas oportunidades de colaboración y más.

### **2.1.5 Beneficios de su aplicación: Mayor Rentabilidad**

La Inteligencia de Negocios proporciona una vista precisa de toda la información importante del negocio, para ayudar a tomar decisiones con mayor rapidez y confianza. En la mayoría de los casos, se pueden obtener los siguientes resultados tangibles:

**Mayor agilidad de negocios:** La Inteligencia de Negocios combina mecanismos de proyección más precisos y capacidades de análisis con herramientas de información e implementación transaccional en tiempo real, que ayudan a la empresa a transformar una idea en algo concreto casi instantáneamente. El resultado es la capacidad de responder a nuevas oportunidades con una rapidez sin precedentes.

**Menores costos de operación:** La Inteligencia de Negocios reduce el tiempo requerido para recolectar información de negocios relevante y luego faculta a los empleados de todos los niveles a realizar consultas, preparar informes e

implementar los resultados, con poco o ningún apoyo de los administradores de la bases de datos y de los especialistas de informática.

***Adquisición de clientes agilizada:*** La Inteligencia de Negocios permite a las empresas identificar con precisión a los clientes potenciales y llevar a cabo campañas efectivas y enfocadas que pueden reducir sustancialmente los costos de mercadeo y publicidad por cliente.

***Mayor lealtad del cliente:*** La Inteligencia de Negocios proporciona un panorama claro de las necesidades del cliente, facilitando así identificar los productos, características y servicios que valoran sus clientes. Así mismo, la Inteligencia de Negocios proporciona herramientas poderosas para una comunicación y un mercadeo personalizado a través de la Web.

***El resultado = mayor rentabilidad:*** La Inteligencia de Negocios coloca a la empresa en posición para generar el mayor beneficio de las unidades de negocios existentes y anticipar nuevas oportunidades. Los sistemas de Inteligencia de Negocios también pueden ayudar a la empresa a reducir los ciclos de desarrollo de productos, agilizar las operaciones, afinar campañas mercadeo y mejorar las relaciones con clientes y proveedores, lo cual conlleva a menores costos y a mayores márgenes de utilidad.

### **2.1.6 Data Warehousing y Data Marts**

El mecanismo más habitual para estructurar la información para un sistema de Inteligencia de Negocios es haciendo uso de un Data Warehouse. Las definiciones más habituales de este término son:

- Almacén de datos. Plataforma que concentra la información de interés de toda la empresa.

- Sistema que permite el almacenamiento en un único entorno de la información histórica e integrada proveniente de los distintos sistemas de la empresa, y que refleja los indicadores clave asociados a los negocios de la misma.
- Sistema de información orientado a la toma de decisiones empresariales que, almacenando de manera integrada la información relevante del negocio, permite la realización de consultas complejas con tiempos de respuesta cortos.
- Sistema orientado a dar información en términos de negocio en vez de datos en términos de explotación.

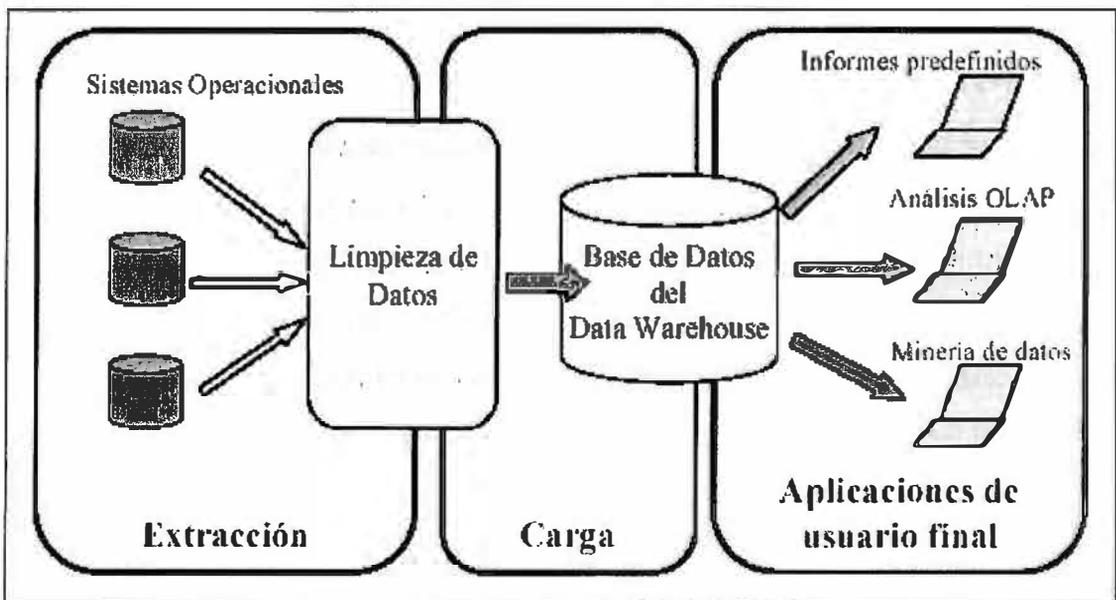
Como se puede apreciar, las palabras más empleadas son: información de interés, negocio, integración, etc. De su conjunto podemos expresar que el Data Warehouse es un almacén estructurado de la información clave del negocio, que integra datos provenientes de todos los departamentos, sistemas, etc., y que nos permite analizar el funcionamiento de la empresa y tomar decisiones sobre su gestión.

No se trata de una simple agregación de las diferentes bases de datos. Es importante destacar que hay algunas diferencias de concepto respecto a éstas y a su forma de uso.

Una base de datos operativa almacena la información de un sector del negocio, se actualiza a medida que llegan datos que deban ser almacenados y se opera mediante los cuatro mecanismos clásicos: “Añadir-Eliminar-Modificar-Imprimir”:

- Clásicamente se orienta hacia la elaboración de informes periódicos.
- Suele manejar pequeños volúmenes de datos.
- Entorno dimensionado para muchas transacciones (gran cantidad de actualizaciones).

En cuanto al Data Warehouse, su actualización se realiza a intervalos regulares (típicamente una al día) dentro de un proceso controlado, y tras realizar un preprocesado de los datos que se van a almacenar. En la figura 2-2 se muestra el proceso de actualización de un Data Warehouse: extracción, carga y explotación por aplicaciones de usuario final.



**Figura 2-2 Proceso de actualización del Data Warehouse**

El Data Warehouse esta orientación hacia la consulta del estado del negocio, dado que:

- Ofrece información bajo demanda (análisis libre mediante el uso de herramientas de generación de informes que atacan el Data Warehouse).
- Refleja el modelo de negocio, frente al modelo de proceso.
- Almacena grandes volúmenes de datos (información histórica e integración de datos de múltiples aplicaciones).
- Esta dimensionado para consultas largas y elaboradas.
- Tiene actualizaciones controladas y no se eliminan datos (el Data Warehouse contiene toda la historia de la compañía).

La estructura de esta gran base de datos es multidimensional, con diferentes puntos de vista que reflejan los distintos aspectos del negocio. Así por ejemplo, los responsables de producto pueden analizar su evolución a lo largo del tiempo en diferentes sectores y localización geográfica. Sobre los mismos datos, los responsables de grandes cuentas pueden obtener información sobre los tipos de productos que se han vendido, por regiones, a lo largo del tiempo. Un director regional podrá estudiar cómo evoluciona su mercado o canal particular, etc.

Cuando mantenemos una estructura de Data Warehouse, pero adaptada sólo a un sector de la empresa, o para un fin concreto, se utiliza un Data Mart. Los Data Marts pueden extraerse del Data Warehouse de la empresa, aunque también es posible que el Data Warehouse se construya a partir de los Data Marts que se hayan ido diseñando e implantando en los diferentes departamentos. Este segundo enfoque es el que se utiliza cuando se comienza por aplicar estas técnicas en algunas de las áreas de la empresa y no en su globalidad.

## ***2.2 Minería de Datos: Herramienta de la Inteligencia de Negocios para desarrollo de Modelos de Estimación y Predicción***

A estas alturas ya va pareciendo claro que si la empresa almacena la información más relevante de su negocio en un sistema de Inteligencia de Negocios que acumula y acumula datos sin parar, un análisis razonable nos puede permitir descubrir tendencias, localizar grupos de datos con comportamiento homogéneo o establecer relaciones valiosas; por ejemplo:

- Que para una gran superficie (supermercados): Más del 60% de las personas que adquieren queso fresco compran algún tipo de mermelada.
- para un departamento de fidelización de una compañía aérea: muchos usuarios que hacen vuelos de menos de 3 días a Berlín alquilan un coche en el aeropuerto.

- para un operador de telefonía: durante el mes siguiente al lanzamiento de una campaña de descuento en llamadas internacionales por parte de una compañía de la competencia, nuestros pequeños clientes redujeron su consumo en este sector, mientras que los grandes clientes lo mantuvieron.

Esta información está oculta en los datos y puede ser extraída haciendo uso de diversas técnicas, como la Minería de Datos, la cual es descrita en esta sección.

### 2.2.1 ¿Qué es Minería de Datos?

La Minería de Datos puede definirse como la **extracción no trivial de información implícita, previamente desconocida y potencialmente útil, a partir de los datos**. Para conseguirlo hace uso de diferentes tecnologías que resuelven problemas típicos de agrupamiento automático, clasificación, asociación de atributos y detección de patrones secuenciales. La Minería de Datos es, en principio, una fase dentro de un proceso global, denominado descubrimiento de conocimiento en bases de datos (Knowledge Discovery in Databases o KDD).

La Minería de Datos permite descubrir patrones significativos y relaciones desconocidas o No evidentes en los datos analizados. Debido a que el análisis visual de las tablas y registros de información generados por la empresa raramente permiten descubrir patrones útiles, los datos son analizados por un proceso automatizado comúnmente denominado “Descubrimiento del Conocimiento”. El proceso de descubrimiento del conocimiento es un componente de la Minería de Datos, que usa la potencia de la computadora, combinada con la habilidad de un operador humano. Por este proceso, la computadora descubre patrones y tendencias presentes en la información analizada, mientras la persona a cargo del análisis decide que patrones son realmente relevantes.

La Minería de Datos puede encontrar patrones descriptivos y predictivos. Cuando se buscan patrones predictivos, el objetivo es derivar información que permita determinar eventos futuros. Por ejemplo, si analizamos información de consumo de reactivos para un proceso químico, asumiendo que contamos con suficiente información para elaborar un modelo válido, el modelo permitirá predecir el consumo exacto para un período futuro definido.

Cuando se buscan patrones descriptivos, el objetivo es describir con precisión eventos pasados y sus causas.

Es habitual que los expertos en estadística confundan la minería de datos con un análisis estadístico de éstos. La diferencia fundamental entre ambas técnicas es muy clara: para conseguir una afirmación como “más del 60% de las personas que adquieren queso fresco compran también algún tipo de mermelada” utilizando un paquete estadístico, es necesario conocer a priori que existe una relación entre el queso fresco y la mermelada, y lo que realizamos con nuestro entorno estadístico es una cuantificación de dicha relación.

En el caso de la Minería de Datos el proceso es muy distinto: la consulta que se realiza a la base de datos busca relaciones entre parejas de productos que son adquiridos por una misma persona en una misma compra. De esta información, el sistema deduce, junto a otras muchas, la afirmación anterior. Como se puede ver, en este proceso se realiza un acto de descubrimiento de conocimiento real, puesto que no es necesario ni siquiera sospechar la existencia de una relación entre estos dos productos para encontrarla.

### **2.2.2 ¿Cuándo aplicamos Minería de Datos?**

La evolución de la tecnología ha facilitado y automatizado en gran medida las tareas de análisis de información. Cada paso en esta evolución se apoya en los anteriores y cada uno de ellos ha supuesto un avance significativo para el usuario

final, que ha visto como cada progreso le abría nuevas posibilidades de análisis y aumentaba el nivel de abstracción de las consultas.

Para decidir cuál es la técnica más adecuada para una determinada situación, es necesario distinguir el tipo de información que se desea extraer de los datos. Según su nivel de abstracción, el conocimiento contenido en los datos puede clasificarse en distintas categorías y requerirá una técnica más o menos avanzada para su recuperación:

➤ **Conocimiento Evidente**

Información fácilmente recuperable con una simple consulta (SQL<sup>4</sup>). Un ejemplo de este tipo de conocimiento es una pregunta como “¿Cuáles fueron las ventas en Lima el mes pasado?” o “¿Cuántos productos fueron defectuosos en el último lote de producción?”.

➤ **Conocimiento multi-dimensional**

El siguiente nivel de abstracción consiste en considerar los datos con una cierta estructura. Por ejemplo, en vez de considerar cada transacción individualmente, las ventas de una compañía pueden organizarse en función del tiempo y de la zona geográfica, y analizarse con diferentes niveles de detalle (país, región, departamento, provincia, distrito).

Técnicamente, se trata de reinterpretar una tabla con  $n$  atributos independientes como un espacio  $n$ -dimensional, lo que permite detectar algunas regularidades difíciles de observar con la representación mono dimensional clásica. Este tipo de información es la que analizan las herramientas OLAP (On-Line Analytical Processing), que resuelven de forma automática cuestiones como “¿Cuáles fueron las ventas de un canal, en una marca y semana específica?”. Por ejemplo, en la figura 2-3 se muestra una

---

<sup>4</sup> Structured Query Language, Lenguaje de consulta directa a bases de datos relacionales.

consulta OLAP que muestra las unidades y el valor venta, así como el costo asociado a las ventas de la marca Coca-cola, en el canal alimentos, en la semana del 24.

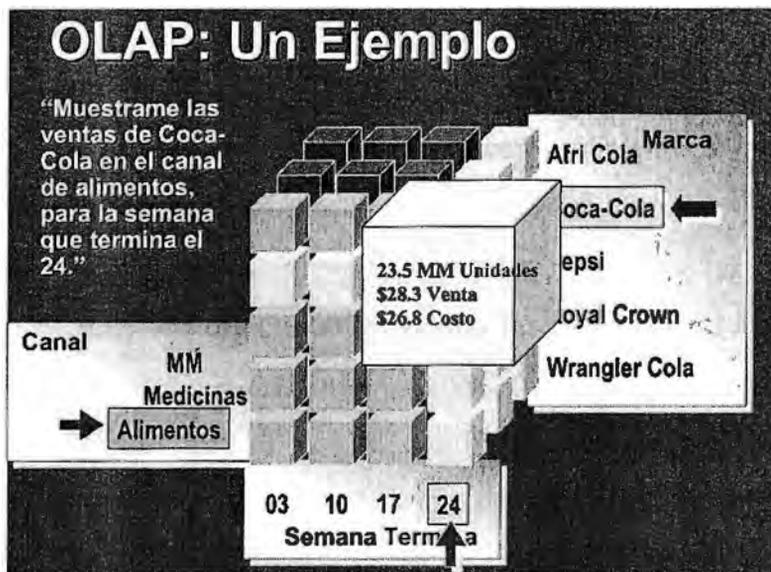


Figura 2-3 Ejemplo de Consulta OLAP

### ➤ Conocimiento oculto

Información no evidente, desconocida a priori y potencialmente útil, que puede recuperarse mediante técnicas de Minería de Datos, como reconocimiento de patrones y relaciones. Esta información es de gran valor, puesto que no se conocía y se trata de un descubrimiento real de nuevo conocimiento, del que antes no se tenía idea, y que abre una nueva visión del problema. Un ejemplo de este tipo sería “¿Qué tipos de clientes tenemos? ¿Cuál es el perfil típico de cada clase de usuario?”.

Como se ve, las técnicas disponibles para extraer la información contenida en los datos son muy variadas y cada una de ellas es complementaria al resto, no exclusivas entre sí. Cada técnica resuelve problemas de determinadas características, y para extraer todo el conocimiento oculto en general, será necesario utilizar una combinación de varias.

La mayor parte de la información de interés contenida en una base de datos, aproximadamente el 80%, corresponde a conocimiento superficial, fácilmente recuperable mediante consultas sencillas (con SQL). El 20% restante corresponde a conocimiento oculto que requiere técnicas más avanzadas de análisis para su recuperación. Estas cifras pueden dar la falsa impresión de que la cantidad de información recuperable mediante técnicas de minería de datos es despreciable. Sin embargo, se trata precisamente de información que puede resultar de vital importancia para la empresa y que no se puede desdeñar.

Básicamente, la clave que diferencia la minería de datos respecto de las técnicas clásicas es que el análisis que realiza es exploratorio, no corroborativo. Se trata de descubrir conocimiento nuevo, no de confirmar o desmentir hipótesis. Con cualquiera de las otras técnicas es necesario tener una idea concreta de lo que se está buscando, y por tanto, la información que se obtiene con ellas está condicionada a la idea preconcebida con que se aborde el problema. Con la minería de datos, es el sistema y no el usuario el que encuentra las hipótesis, además de comprobar su validez.

La minería de datos, esencialmente, permite obtener a partir de los datos un modelo del problema que se analiza, bien sean las ventas de un artículo para mejorar la campaña de marketing, las características técnicas de un producto en control de calidad, o un proceso industrial cuyo control se desea optimizar, por citar algunos ejemplos. El modelo obtenido permitirá simular el comportamiento del sistema real y obtener conclusiones aplicables en el día a día.

### 2.2.3 Minería de Datos frente a OLAP y DSS

En la figura 2-4 se muestran comparativamente las diferentes herramientas de análisis existentes en función al potencial de análisis que ofrecen y a la complejidad del proceso requerido.

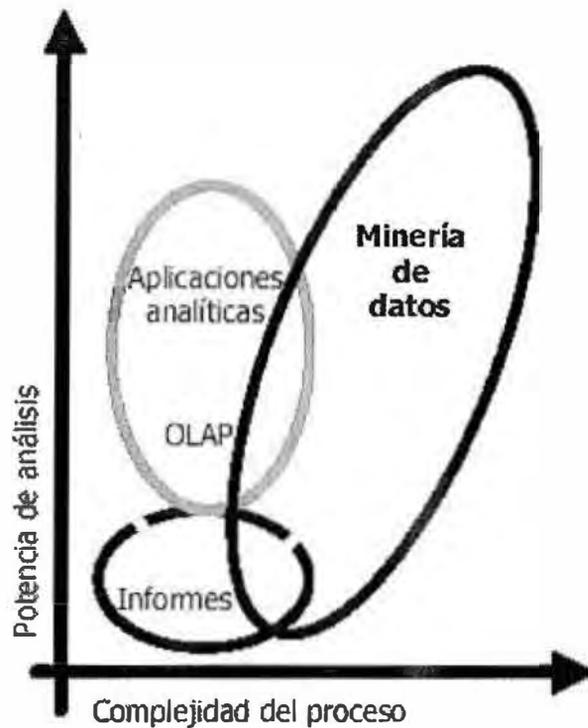


Figura 2-4 Comparación de Herramientas de análisis

Los sistemas de soporte a la toma de decisiones (DSS<sup>5</sup>) son aplicaciones analíticas sobre las que se apoyan los responsables de una empresa, directivos y gestores, en la toma de decisiones. Para ello utilizan:

- Un Data Warehouse, en el que se almacena la información de interés para la empresa, y
- Herramientas de análisis multidimensional (OLAP)

---

<sup>5</sup> Decision Support System

OLAP (On-Line Analytical Processing) se define como análisis rápido de información multidimensional compartida. El término OLAP aparece en contraposición al concepto tradicional OLTP (On-Line Transactional Processing), que designa el procesamiento operacional de los datos, orientado a conseguir la máxima eficacia y rapidez en las transacciones (actualizaciones) individuales de los datos, y no a su análisis de forma agregada.

Las herramientas OLAP permiten navegar a través de los datos almacenados en el Data Warehouse y analizarlos dinámicamente desde una perspectiva multidimensional, es decir, considerando unas variables en relación con otras y no de forma independiente entre sí, y permitiendo enfocar el análisis desde distintos puntos de vista. Esta visión multidimensional de los datos puede visualizarse como un “cubo de Rubik”, que puede girarse para examinarlo desde distintos puntos de vista, y del que se pueden seleccionar distintas “rodajas” o “cubos” dependiendo de los aspectos de interés para el análisis.

Los DSS permiten al responsable de la toma de decisiones consultar y utilizar de manera rápida y económica las enormes cantidades de datos operaciones y de mercado que se generan en una empresa. Gracias al análisis OLAP, pueden verificarse hipótesis y resolverse consultas complejas. Además, en el curso del análisis, la interpretación de los datos puede dar lugar a nuevas ideas y enfoques del problema, sugiriendo nuevas posibilidades de análisis.

Sin embargo, el análisis OLAP depende de un usuario que plantee una consulta o hipótesis. Es el usuario el que lo dirige y, por tanto, el análisis queda limitado por las ideas preconcebidas que aquél pueda tener.

La minería de datos constituye un paso más adelante en el análisis de los datos de la empresa para apoyar la toma de decisiones. No se trata de una técnica que sustituya los DSS ni el análisis OLAP sino que los complementa, permitiendo realizar un análisis más avanzado de los datos y extraer más información de ellos.

Como ya se ha comentado en la sección anterior, al utilizar la minería de datos es el propio sistema el que descubre nuevas hipótesis y relaciones. De este modo, el conocimiento obtenido con estas técnicas no queda limitado por la visión que el usuario tiene del problema.

Las diferencias entre minería de datos y OLAP radican esencialmente en que el enfoque desde el que se aborda el análisis con cada una de ellas es completamente distinto. Fundamentalmente:

El análisis que realizan las herramientas OLAP es dirigido por el usuario, deductivo, parte de una hipótesis o de una pregunta del usuario y se analizan los datos para resolver esa consulta concreta. Por el contrario, la minería de datos permite razonar de forma inductiva a partir de los datos, para llegar a una hipótesis general que modele el problema.

Además, las aplicaciones OLAP trabajan generalmente con datos agregados, para obtener una visión global del negocio. Por el contrario, la minería de datos trabaja con datos individuales, concretos, descubriendo las regularidades y patrones que presentan entre sí y generalizando a partir de ellos.

Así, en la tabla 2-1 se muestra una comparación en el razonamiento y forma de trabajo de las herramientas OLAP versus las de minería de datos.

	<b>OLAP</b>	<b>Minería de Datos</b>
Razonamiento	Deductivo	Inductivo
Trabaja con datos	Agregados	Concretos/individuales

**Tabla 2-1 Razonamiento y forma de trabajo de herramientas OLAP y Minería de Datos**

Un ejemplo clarificará la diferencia entre ambas técnicas:

Una pregunta típica de un sistema OLAP/DSS sería: “El año pasado, ¿se compraron más motos en Loreto o en Lima?”. La respuesta del sistema sería del tipo “En Loreto se compraron 12,000 motos, mientras que durante el mismo intervalo en Lima se compraron 15,000 motos”. Obviamente es una información interesante y útil, pero restringida por las hipótesis realizadas a priori.

En cambio, un problema típico para resolver utilizando minería de datos sería, por ejemplo: “Hallar un modelo que determine las características más relevantes de las personas que compraron motos”. A partir de los datos del pasado, el sistema de minería de datos proporcionaría una respuesta del tipo: “Depende de la época del año y la situación geográfica. En invierno, los habitantes de Lima que pertenecen a un cierto grupo de edad y nivel de ingresos probablemente comprarán más motos que gente de las mismas características en Loreto”.

Como puede verse, se trata de problemas distintos, de modo que según los objetivos perseguidos deberá utilizarse una técnica u otra. Además, puesto que sus conclusiones son complementarias, en general será conveniente combinar ambas para obtener los mejores resultados.

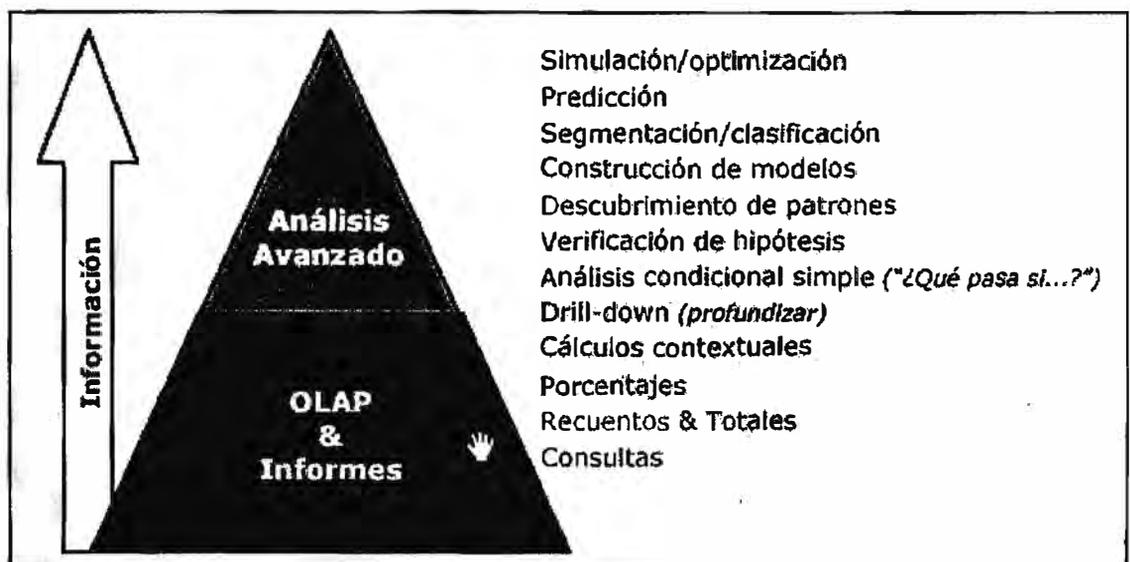


Figura 2-5 Herramientas disponibles según problema a resolver

Como se muestra en la figura 2-5, la herramienta de análisis a emplear dependerá del problema que se desee resolver, el cual puede variar de consultas simples a predicciones o simulaciones.

#### **2.2.4 Usos frecuentes de la Minería de Datos**

El objetivo final de cualquier proyecto de minería de datos puede resumirse en uno de esto dos:

- Ahorrar dinero mejorando la eficacia de sus actividades, o bien,
- Ganar dinero descubriendo nuevas fuentes de beneficios

¿Cómo se llega a estos objetivos? A partir de un conjunto de datos y un conjunto de técnicas se puede llegar a unas determinadas conclusiones. Pero, ¿Cómo se traducen los resultados de un proyecto de minería de datos en beneficios tangibles para la empresa? Básicamente, esos resultados suponen una mejora de la información disponible y será al aplicar dicha información cuando se obtengan los beneficios.

Los campos en los que pueden utilizarse estas técnicas son extremadamente variados: prácticamente en cualquier situación en la que se disponga de un conjunto de datos. A continuación, se comentan algunas de las áreas más comunes en las que se ha aplicado frecuentemente la minería de datos, pero se trata simplemente de algunos ejemplos.

A modo de curiosidad: 28 de 29 equipos que participan en la liga de baloncesto profesional americana (NBA) utilizan técnicas de minería de datos para detectar patrones de comportamiento y relaciones entre variables del juego (por ejemplo, detectar que el jugador X realiza el 90% de sus tiros de campo cuando el jugador Y juega de base), de forma que estas técnicas ofrecen nuevas perspectivas para modificar las tácticas de juego a fin de mejorar el rendimiento del equipo. Un

análisis tradicional podría indicar que un jugador consigue el 70% de sus puntos en tiros de media distancia desde el lateral derecho.

En general, disponer de un modelo que permita simular el comportamiento y/o predecir la evolución de un sistema, un proceso, las ventas de un producto, etc., de forma suficientemente precisa, supone una clara ventaja competitiva, permitiendo adelantarse y aprovechar oportunidades, así como prevenir problemas. Algunas de las aplicaciones más comunes de la minería de datos son las siguientes:

### ***Marketing***

Este es uno de los campos donde los éxitos de la minería de datos son más conocidos. Cuanto más precisa sea la información que tengamos sobre los clientes, mayores posibilidades tendremos de aumentar nuestros ingresos y rentabilizar al máximo nuestras acciones. El objetivo fundamental puede resumirse en determinar quién comprará qué, cuándo y dónde.

- Targeting: Podemos aumentar espectacularmente el porcentaje de respuesta a una campaña de marketing si se dirige a los objetivos adecuados. La minería de datos permite detectar entre los potenciales clientes los que presentan una mayor probabilidad de responder a la campaña y dirigirla a ellos específicamente, con lo cual se consigue reducir drásticamente los costos.
- Fidelización de clientes: Conseguir un nuevo cliente o recuperar uno perdido resulta mucho más costoso que mantener uno que ya lo es. De ahí la rentabilidad de las campañas de fidelización de clientes, que detectan aquellos que parece más probable que se vayan a perder, permitiendo llevar a cabo iniciativas que eviten dicha pérdida.
- La minería de datos también permite detectar nuevas oportunidades de mercado, comparando hábitos de consumo de diferentes clientes, por

ejemplo, o determinando la ubicación más conveniente para un determinado negocio.

### ***Predicción***

Conocer a priori cómo evolucionará una variable en el futuro constituye una información muy valiosa y supone una indudable ventaja competitiva. Se trata de una herramienta de evidente interés tanto desde el punto de vista comercial, como en gestión o control de procesos.

A partir de los datos históricos almacenados y utilizando técnicas de minería de datos pueden elaborarse modelos que permitan estimar con precisión la evolución de una variable en el futuro. Disponer de esta información con tiempo suficiente permite adecuar la respuesta de forma óptima. Esto puede resultar útil en los campos de:

- Detección de oportunidades
- Prevención de problemas
- Gestión óptima de personal
- Optimización de stocks

### ***Reducción de Riesgos***

La minería de datos permite construir sistemas de evaluación automática de riesgos, basados en la experiencia previa. Estos sistemas resultan de gran utilidad cuando la cantidad de casos a evaluar es excesiva para su procesamiento manual. El empleo de técnicas de minería de datos ha aumentado la eficacia y fiabilidad de dichos sistemas, logrando un comportamiento más similar al de los expertos humanos.

### ***Detección de fraudes***

Aplicando técnicas de minería de datos, pueden obtenerse modelos que permitan descubrir posibles fraudes, basándose en la detección de comportamientos anómalos, en comparación con los datos registrados anteriormente.

Podemos encontrar aplicaciones concretas en operadores de telefonía o empresas de gestión de tarjetas de crédito. Estas compañías analizan el uso que los clientes hacen de sus servicios y pueden localizar, de manera muy rápida, un uso fraudulento de los mismos.

### ***Control de Calidad***

Existen numerosos ejemplos en los que se han aplicado técnicas de minería de datos para desarrollar sistemas automáticos de control de calidad. Estos sistemas suponen un considerable ahorro en el proceso productivo, puesto que facilitan:

#### **➤ Detección más precisa de productos defectuosos**

A menudo el control de calidad se realiza de forma manual y, por tanto, depende de una evaluación subjetiva por parte del personal encargado del mismo. El principal problema de este método es que el criterio de calidad no es estable sino que depende de la persona que realiza el análisis. La minería de datos permite desarrollar sistemas automáticos de control de calidad, que discriminan los productos defectuosos con un alto grado de precisión y fiabilidad, según un criterio objetivo.

Esto no sólo evita el problema mencionado anteriormente. Además, al aumentar la exactitud de la evaluación, se ahorran los costos derivados de las clasificaciones erróneas: productos defectuosos que se consideraron correctos por error y productos correctos, desechados por un exceso de precaución.

➤ **Localización precoz de defectos**

El control de calidad no sólo debe realizarse al final del proceso. Cuanto antes se detecte un fallo, menor será su impacto. Además de las ventajas de los sistemas automáticos ya comentados, en este caso existe un problema añadido. A menudo no resulta fácil medir la variable que determina la calidad del producto en tiempo real o en la cadena de producción. En estos casos, es imprescindible utilizar técnicas de minería de datos para descubrir posibles relaciones que permitan detectar los fallos, utilizando las variables disponibles durante el proceso.

➤ **Identificación de causas de fallas**

La minería de datos no sólo resulta útil para discriminar los productos defectuosos. También ayuda a determinar las fallas más frecuentes, así como identificar las causas de los mismos. Esto permite adoptar medidas para evitarlos en el futuro.

➤ **Análisis no destructivo**

A menudo, para obtener la información que se necesita, hay que realizar un análisis destructivo. Un ejemplo típico es la evaluación de la resistencia de un material, medida que se establece forzándolo hasta que se rompe. Utilizando minería de datos es posible estimar con bastante exactitud el valor de este tipo de parámetros, en función de otras características que sí pueden medirse sin destruir el producto. Esto permite controlar la calidad de todos los productos fabricados y no sólo de una pequeña muestra, ya que no se destruyen con el examen.

### ***Procesos Industriales***

Otra aplicación básica de la minería de datos en el entorno industrial, además del control de calidad, es el control de procesos. Estas técnicas permiten explotar la información disponible sobre un sistema o proceso y utilizar los modelos

desarrollados (bien de un sistema o proceso global, o bien de una parte concreta del mismo), para:

➤ **Automatizar y optimizar el control del proceso**

En muchos sistemas se conoce el proceso suficientemente como para diseñar e implantar controladores a partir de análisis matemático del proceso. En otras ocasiones, esto no es posible, bien por que el proceso es enormemente complejo, bien porque no disponemos de todas las variables. En estas circunstancias, técnicas de minería de datos pueden ayudarnos a establecer relaciones entre las variables, y así diseñar los controladores adecuados.

➤ **Optimizar su rendimiento**

Los propios sistemas de aprendizaje pueden ser utilizados para adaptar los mecanismos de control de forma permanente, en función de los datos del proceso que vayamos recibiendo. De esta forma, es posible optimizar el rendimiento del proceso, adaptando los controladores, en cada momento, a la situación de la planta.

➤ **Implementar programas de mantenimiento predictivo**

Uno de los problemas de todo equipo de mantenimiento de un proceso es establecer el calendario de reparaciones. Las reparaciones, limpiezas y ajustes programados suponen en muchos casos parar el proceso productivo, con las consiguientes pérdidas, no sólo de lo que se deja de producir sino de los costos de parada y arranque de la cadena. Un análisis profundo de los datos de que se disponga puede permitir hacer una planificación óptima de estas paradas, de manera que se minimice su impacto.

## 2.2.5 Fases de un proyecto de Minería de Datos

Como toda disciplina relacionada con sistemas de información, la Minería de Datos requiere el desarrollo y seguimiento de un plan de trabajo, basado en fases. Se inicia desde la concepción de la idea del proyecto hasta su implementación. En la figura 2-11 se muestra gráficamente las fases de un Proyecto de Minería de Datos:

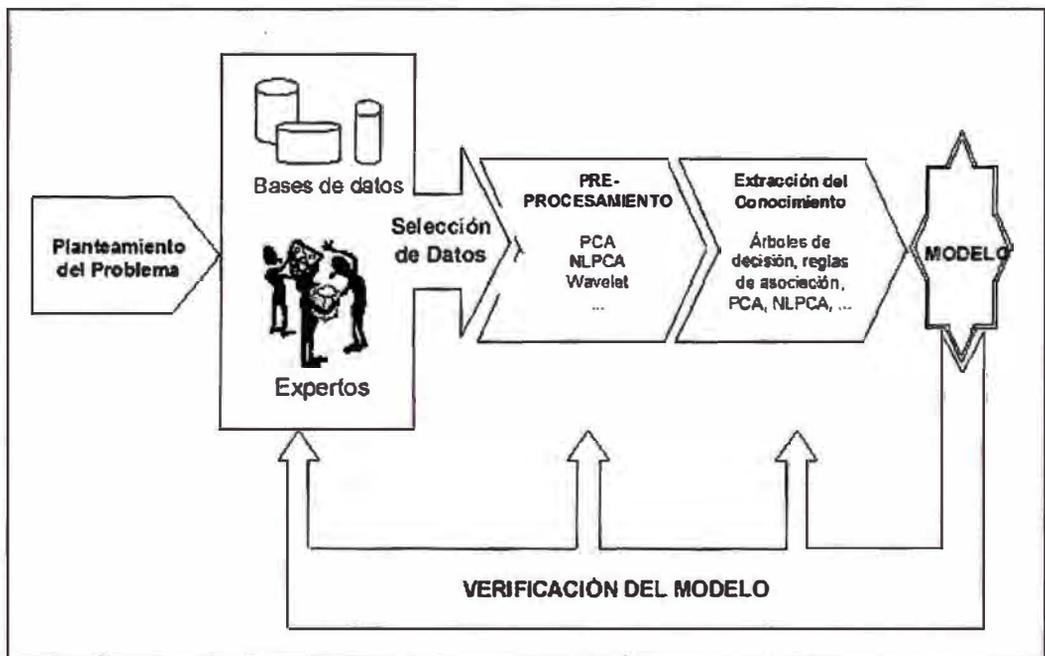


Figura 2-6 Fases de un proyecto de Minería de Datos

Estas fases son:

- Planteamiento del problema
- Selección de datos
- Preprocesamiento de datos
- Extracción del conocimiento
- Construcción del modelo
- Validación del modelo

### ***Planteamiento del problema***

En esta primera etapa del proyecto se debe definir de manera objetiva cuál es el problema a resolver, determinar con qué recursos humanos y tecnológicos se cuenta, cuales son las fuentes de información y que disponibilidad de información se tiene.

### ***Selección de los datos***

La fuente de datos debe ser evaluada y analizada para determinar cual de ellas puede ser candidata para aplicar criterios de Minería de Datos. La calidad y abundancia de datos son los factores principales para la elección de la base de datos de información a analizar.

Se debe realizar un tratamiento y estructuración de la información con el objetivo de presentarla de la mejor manera posible para su análisis posterior.

Los datos son extraídos de una fuente nativa, es decir de los sistemas de transacciones donde residen, archivos textos, hojas de cálculo o base de datos. Estos datos, luego de extraídos son cargados en una base de datos con una estructura compatible con el modelo de Minería a desarrollar. Típicamente, se emplean los Servicios de Transformación de Datos (DTS), para extraer y depurar los datos que no cumplan con los criterios de selección.

### ***Preprocesamiento de datos***

Una vez que los datos han sido extraídos, se analizan con la finalidad de reorganizar la información, eliminando aquella que es de poca utilidad.

### ***Extracción del conocimiento***

En esta fase se escogen y se aplican las técnicas de minería de datos para la determinación de patrones de interés en los datos. Para ello, se interpretan los resultados obtenidos a lo largo del proceso para la construcción de modelos o se buscan estructuras subyacentes dentro de la información. Entre las técnicas de minería de datos se encuentran: árboles de decisión, redes neuronales, reglas de asociación, regresión lineal y no lineal de predicción y clasificación (que a su vez puede involucrar análisis estadístico, redes neuronales, lógica difusa, etc.), algoritmos genéticos, etc.

### ***Construcción del modelo***

Con los resultados obtenidos en la fase anterior se lleva a cabo el análisis, interpretación y evaluación para la determinación de un modelo eficiente que sea útil en la toma de decisiones.

### ***Validación del modelo***

En esta fase se implementa el modelo desarrollado en el proceso real y se determina su efectividad en diferentes casos de aplicación. Si las pruebas arrojan resultados satisfactorios el modelo queda comprobado y garantizado para su uso regular. Sin embargo, si los resultados son insatisfactorios, se deberá regresar a analizar las fases anteriores y fortalecer para mejorar el modelo final.

## **2.2.6 Aprendiendo de los datos históricos**

Si bien la minería de datos permite predecir resultados futuros, también permite sacar conclusiones acerca de los eventos pasados. Cuando se buscan conclusiones de eventos pasados, el objetivo de la minería de datos no es la predicción sino el entendimiento de la relación causa-efecto entre los elementos de datos analizados.

Existen seis formas de usar la minería de datos para ganar conocimiento sobre eventos pasados.

### ***Análisis de influencia***

Un análisis de influencia determina como los factores y variables analizados impactan en una forma importante en otras variables, como puede ser la calidad de un proceso. Por ejemplo, un gerente de un centro de ventas por teléfono que se encarga de colocar tarjetas de crédito desea conocer que factores contribuyen a que la llamada pueda colocar una tarjeta. Existen muchas variables que pueden afectar el éxito de la llamada, tales como:

- Frecuencia de llamada
- Tiempo empleado por llamada
- Tipo de llamada – entrante o saliente
- Sexo de la persona contactada
- Sexo de la persona que realiza la llamada
- Hora del día en que la llamada fue hecha

El análisis de influencia le permitirá al gerente entender el ambiente y las condiciones que afectan el éxito de sus llamadas telefónicas. También en base a este análisis podrá definir las políticas de operación del centro telefónico. Este modelo no es predictivo, simplemente permite entender la relación causa-efecto escondida en los datos históricos.

### ***Análisis de Variación***

Un análisis de variación busca las variaciones en un conjunto de datos y permite aislar algún factor que influencia en gran medida a estas variaciones. Encontrar variaciones en los datos es importante cuando se necesita descubrir que factores ocasionan estas diferencias. Por ejemplo, esto puede ser útil para revisar las

diferencias en la eficiencia de llamadas por localidad, por hora del día y por gerencia. Un análisis de variación puede permitir realizar la siguiente conclusión: “Las centrales telefónicas de venta en grandes ciudades tienen menor eficiencia que las centrales en pequeños pueblos”

### ***Análisis Comparativo***

Usando el ejemplo anterior, un análisis comparativo puede ser usado para comparar la eficiencia de las llamadas hechas en dos centros telefónicos en dos diferentes localidades para decidir cual de ellos puede ser ampliado y cual cerrado.

### ***Análisis causa-efecto***

Un análisis causa-efecto determina el efecto de un evento dado. Por ejemplo, es evidente que un incremento del 50% en las llamadas por hora de un centro telefónico ocasionará una reducción severa en la eficiencia de las llamadas, pero es menos evidente detectar que las llamadas fraudulentas son reportadas por operadores en su sobre tiempo, cuando las llamadas reportadas se incrementan en casi la mitad. Un análisis causa-efecto podría revelar que el incremento en llamadas realizadas también incrementa las pérdidas debido a fraude.

### ***Análisis de Tendencias***

Un análisis de tendencias busca cambios en el valor de una variable sobre un período específico de tiempo. Los datos son medidos en términos de dirección de movimiento en lugar de un valor cuantitativo. Por ejemplo, las tendencias usadas para describir el estado del mercado de valores, como cuando se dice “el mercado está en alza” o “el mercado está a la baja”, estas frases expresan la diferencia entre el valor actual y el valor previo del mercado.

### ***Análisis de Desviaciones***

Un análisis de desviaciones identifica datos que están fuera de su rango normal esperado. La detección de fraudes emplea el análisis de desviaciones para detectar valores sospechosos. Por ejemplo, una compañía de tarjetas de crédito puede detectar en línea que una tarjeta ha sido robada, si los consumos realizados son fuera de los normalmente realizados por la dueña de la persona.

#### **2.2.7 Prediciendo el futuro**

Se puede predecir el futuro aplicando los patrones descubiertos por el motor de minería de datos para la construcción de un modelo predictivo. A diferencia del modelo descriptivo, el cual es diseñado para encontrar patrones y ayudas para entender los datos analizados; el modelo predictivo usa el patrón descubierto para realizar el mejor estimado de los valores de salida para un nuevo conjunto de datos de entrada.

Es importante determinar la probabilidad de éxito del modelo predictivo, es decir el resultado del modelo predictivo debe estar asociado a su probabilidad.

### 2.3 Transformando la Planta en un almacén de información

Los factores sociales y financieros que manejan el negocio industrial están cambiando. Para permanecer competitivos, los sistemas de producción deben ir cambiando con las necesidades del mercado, las nuevas directrices del negocio y la competencia.

La figura 2-12 presenta las necesidades actuales del mercado:

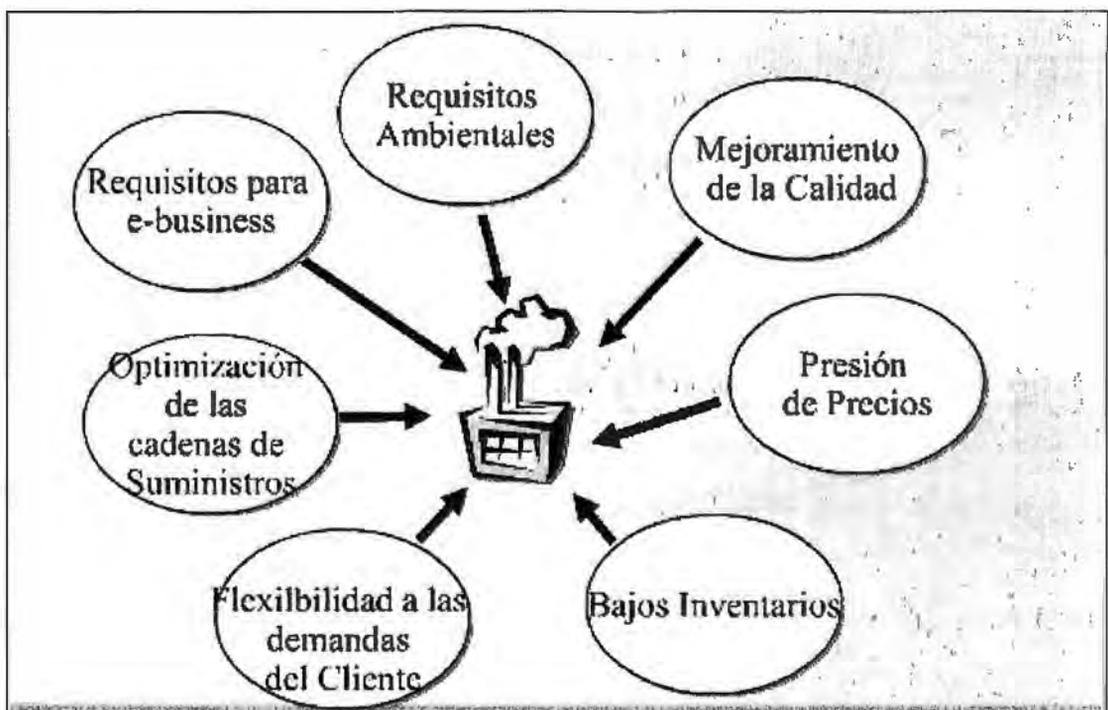


Figura 2-7 Necesidades del mercado

Las nuevas directrices de negocios exigen que:

- Los procesos manufactureros estén distribuidos “donde convienen” y “cuando” sean requeridos.
- Un proceso manufacturero se desarrolla en la planta, pero se extiende más allá de los bordes de la empresa.

- Las plantas no sean consideradas “entidades separadas”, sino que se ubiquen dentro de la cadena de suministros: proveedor-empresa-cliente

Y la competencia está cambiando las reglas:

- La interacción con los clientes es más cercana y no se limita a ordenes y entrega de pedidos
- La reducción del tamaño de lotes, tiempos de entrega más pequeños, y administración de cambios en los pedidos, son elementos claves para el éxito de la empresa.
- El cumplimiento de las reglas y las regulaciones representan una significativa parte de los costos productivos.

Por estas razones, la información generada en planta se convierte en un insumo importante aún no explotado, para generar valor en la empresa.

### **2.3.1 El valor de la información de planta para la empresa**

Con mejor información se podrán tomar mejores decisiones en el área de manufactura. Estas decisiones están orientadas a conseguir:

- Menor tiempo de llegada al mercado (time to market), lo que permite una ventaja competitiva
- Reducción de costos, lo que permite mejorar la rentabilidad de la empresa e incrementar la participación de mercado
- Reducción de los tiempos de paro comprendiendo las causas reales de falla, lo que permite incrementar la productividad de los activos
- Optimizar tiempos de proceso, lo que permite producir más con los mismos recursos, es decir incrementar la productividad de la empresa
- Optimizar operaciones, lo que permite reducir inventarios de materia prima, material en proceso y productos terminados, con lo que reducimos el capital inmovilizado por inventarios.

Al utilizar de una manera eficiente los recursos de planta, los fabricantes pueden ser más productivos con el equipo existente, generando mejores resultados económicos a la organización. El valor de la información generada en planta reside en que su correcto uso incide directamente en la mejora de los resultados de la empresa y esto genera **Rentabilidad**.

### 2.3.2 ¿Cómo agrega valor a la empresa la información de manufactura?

Si analizamos la figura 2-14, la cadena de valor en manufactura:

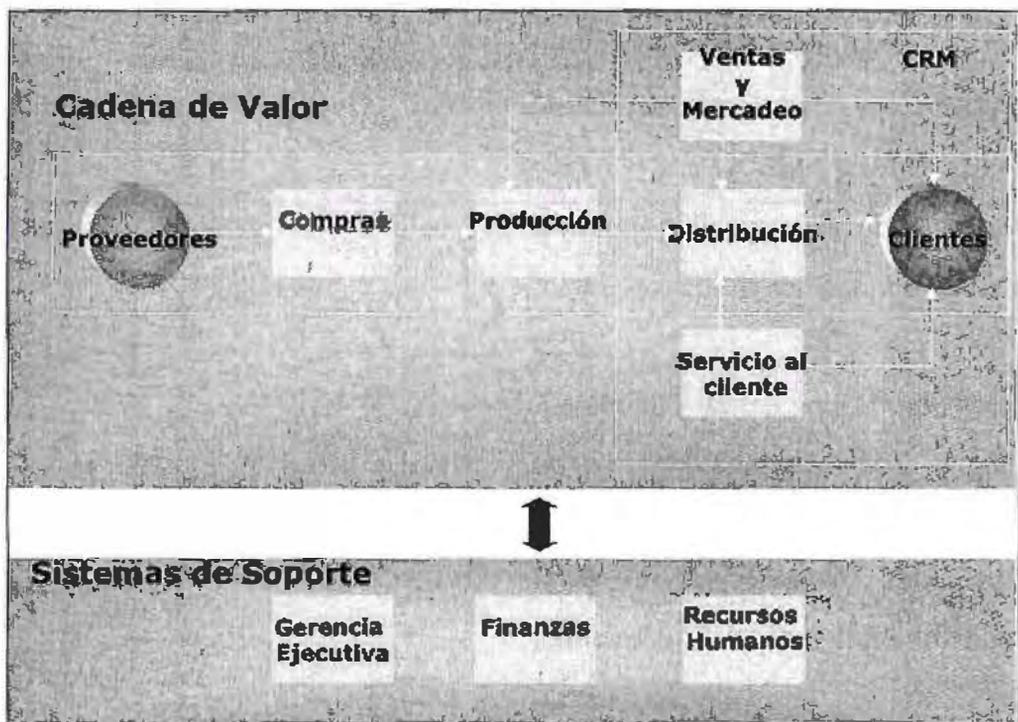


Figura 2-8 Cadena de valor en manufactura

Podemos determinar que la información de manufactura puede impulsar acciones que generen mejoras significativas en tres áreas básicas de la empresa: Ventas y Mercadeo, Producción y Finanzas. La tabla 2-2 presenta las mejoras alcanzables y las acciones impulsoras para estas mejoras.

Área	Mejoras alcanzables	Acción impulsora
Ventas y Mercadeo	Incremento de ventas Mayor participación de mercado	<ul style="list-style-type: none"> <li>• Menor tiempo de llegada al mercado (time to market)</li> </ul>
Producción	Calidad y Rendimiento	<ul style="list-style-type: none"> <li>• Reducción de los tiempos de paro, comprendiendo las causas reales de falla, lo que permite incrementar la productividad de los activos</li> <li>• Incrementar el throughput, lo que permite producir más con los mismos recursos, es decir incrementar la productividad de la empresa</li> <li>• Optimizar parámetros de operaciones para elevar calidad de productos</li> </ul>
Finanzas	Mejorar rentabilidad de empresa	<ul style="list-style-type: none"> <li>• Reducción de costos de operación</li> <li>• Incremento de Eficiencia de los activos fijos</li> <li>• Incremento de Eficiencia del capital de trabajo, mediante reducción inventarios de materia prima, material en proceso y productos terminados, con lo que reducimos el capital inmovilizado por inventarios.</li> </ul>

**Tabla 2-2 Áreas a mejorar en la empresa**

### **2.3.3 La integración entre la planta y la administración de la empresa: Los sistemas MES (Manufacturing Execution Systems)**

La capacidad para obtener información de los dispositivos de planta y convertir los datos crudos en información útil es el principio central detrás de los sistemas integrados de manufactura.

El reto de la integración está en integrar no sólo los diferentes sistemas y dispositivos de control, sino también las diferentes áreas de la planta, en tres dimensiones diferentes:

- Dentro de las operaciones
- Entre las operaciones: sistema coordinado de manufactura
- Entre el piso de planta y el resto de la compañía

Como solución a esta necesidad de integración han aparecido en el mercado los sistemas MES (Manufacturing Execution Systems).

Los sistemas MES se utilizan como soporte en la optimización de procesos industriales, dado que permiten integrar información de la planta con el resto de la compañía. Los MES entregan en tiempo real información consolidada y analizada de las operaciones desde el piso de planta a los gerentes o analistas responsables. Esta información permite tomar decisiones sobre asignación de recursos, control de calidad, desempeño y análisis gerencial.

En la figura 2-15 se presentan los diferentes sistemas de información de una empresa productiva, y como el MES permite integrar la información básica de control con los sistemas de planeamiento de recursos.

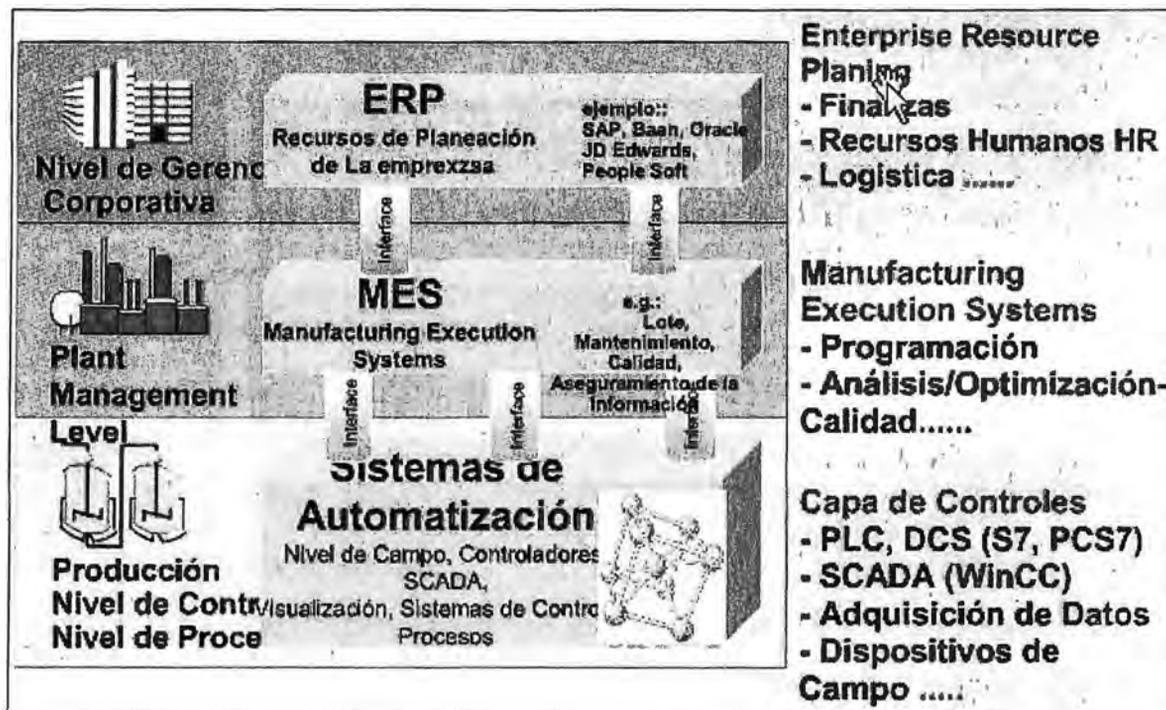


Figura 2-9 Ubicación del MES en una empresa productiva

## 2.4 Aplicación de Minería de datos en los Procesos de Planta

La aplicación de técnicas de Minería de Datos a la información de planta, centralizada por el sistema MES, permite optimizar las operaciones de manufactura. En esta sección se presentan cuatro áreas de alto impacto dentro del proceso de manufactura, donde se ha aplicado con éxito la minería de datos para la optimización de operaciones:

- Administración de flujo de órdenes
- Administración de flujo de producción y materiales
- Administración de los recursos de la empresa
- Sincronización de la cadena de valor

### **2.4.1 Administración de flujo de órdenes**

La aplicación de minería de datos en la administración de flujo de órdenes de trabajo permite optimizar el cumplimiento de la demanda dentro de la planta, integrando los sistemas de planeamiento y transaccionales. Las órdenes son ligadas directamente a sistemas de planeación de capacidad finita y a los sistemas de reporte para producción en el piso de planta.

#### **Aplicación:**

##### ***Seguimiento de órdenes***

- ✓ Los datos capturados en el piso de planta ofrecen información en tiempo real sobre el estado de cualquier orden en particular. Estos datos permiten generar un modelo para optimizar ciclos de proceso.
- ✓ Los datos de costo real – no estándar- son capturados en toda la planta a través de los sistemas del piso de planta. Estos datos permiten generar un modelo para predecir costos de nuevos productos.

##### ***Programación***

- ✓ Los programas de producción se reciben del sistema de planeación empresarial y son optimizados para el piso de planta utilizando planeadores de capacidad finita. Estos planeadores están basados en modelos de minería de datos predictivos.

#### **Impacto en las operaciones:**

Reducción en el inventario de materiales en proceso, reducción en el tiempo de ciclo de operación, mejora en la visibilidad de las órdenes de trabajo, mejoras en la programación de producción.

## **2.4.2 Administración de flujo de producción y materiales**

La aplicación de minería de datos en la administración de flujo de producción y materiales permite optimizar el flujo de materiales y actividades que agregan valor dentro de la planta, así como su sincronización con la cadena de valor. La mejora continua se alcanza utilizando metodologías como Lean Manufacturing, Six Sigma y la instalación de sistemas de información.

### **Aplicación:**

#### ***Eficiencia y Efectividad de Producción***

- ✓ Visualización en tiempo real del estado actual de la producción y su desempeño contra el plan
- ✓ Simulación del flujo de producción

#### ***Lean Manufacturing***

- ✓ Implementación de los cambios de proceso para reducir la variabilidad

#### ***Administración del Inventario y seguimiento de materiales (Tracking)***

- ✓ Registro de la utilización de materias primas para seguimiento de inventario.
- ✓ Simulación de los flujos de materiales

#### ***Administración de la calidad***

- ✓ Registro de los datos de calidad, desde la operación hasta la producción de desperdicios. Esta información permite generar un modelo para determinar parámetros óptimos de operación para elevar calidad de productos y disminuir desperdicios.

***Desempeño Humano***

- ✓ Entrenamiento y cambios operacionales para mejorar la eficiencia y efectividad de las operaciones, basadas en las mejores prácticas de operación determinadas por el modelo de minería de datos.

**Impacto en las operaciones:**

Reducción en las materias primas, en el inventario de materiales en proceso, reducción en el tiempo de ciclo de operación, incremento en la agilidad de planta, reducción de desperdicios y mejora de la calidad de productos.

**2.4.3 Administración de los recursos de la empresa**

La aplicación de minería de datos en la administración de recursos de la empresa permite optimizar el desempeño de los recursos en el piso de planta. Se enfoca en los componentes principales de la operación incluyendo mantenimiento, operación de equipos, desempeño de operadores y monitoreo.

**Aplicación:*****Eficiencia y Efectividad en el mantenimiento***

- ✓ Ofrecer un método para dar seguimiento a las actividades de mantenimiento y determinar causa de fallas y efectos causado por el mantenimiento
- ✓ Implementar sistemas que permitan realizar mantenimiento preventivo y predictivo.

***Confiabilidad de Equipos***

- ✓ Identificar condiciones que resulten en la óptima operación de los equipos.

- ✓ Permitir un monitoreo confiable de la eficiencia operacional de los equipos e identificar las causas que generan un desempeño NO adecuado.

#### ***Tablero de análisis de desempeño***

- ✓ Establecer parámetros de desempeño, implementar tableros para el monitoreo de recursos y actividades de mantenimiento. Identificar oportunidades de mejora utilizando análisis basado en datos reales.

#### **Impacto en las operaciones:**

Reducción de costos de mantenimiento y mejora en la productividad de los recursos de planta.

### **2.4.4 Sincronización de la cadena de valor**

Esta área de aplicación representa la columna vertebral para la coordinación y colaboración entre las operaciones de la planta, los proveedores y clientes. Ofrece un enlace que permite la optimización de procesos de planta a través de la sincronización de información con otras áreas de la empresa y su modelamiento.

#### **Aplicación:**

##### ***Sincronización de la demanda***

- ✓ Relaciona la demanda externa con la de compañía y soporta los sistemas para el soporte de decisiones. El modelar la demanda de la compañía permite modelar y planificar el plan de producción.

##### ***Sincronización con Inventarios***

- ✓ Mejora el flujo de inventarios desde los proveedores hasta las actividades de producción. El modelo de pedido se basa en los tiempos

de entrega y disponibilidades históricas, así como en la demanda futura de la compañía, para estimar momento y cantidad de reposición.

#### ***Sincronización de la Producción***

- ✓ Optimización de la producción mejorando la exactitud del programa de producción y la coordinación del programa maestro de producción con las actividades programadas en el piso de planta.
- ✓ Coordinación de recetas y listas maestras de materiales
- ✓ Simulación de la demanda, inventario y producción

#### ***Sincronización con las órdenes de materiales de reposición***

- ✓ Se enlaza con los requerimientos de materiales de reposición hacia compras y proveedores.

#### ***Análisis del tablero de la cadena de valor***

- ✓ Determinar los parámetros de desempeño y establecer un tablero para monitorear la sincronización con la empresa y la cadena de valor, para programar mejoras basadas en datos reales.

#### **Impacto en las operaciones:**

Mejoras en la optimización de los recursos y las actividades de valor agregado en la empresa a nivel de componentes de la cadena de valor.

### 2.4.5 Otras aplicaciones de la Minería de Datos

En la tabla 2-3 se resumen otras aplicaciones de la minería de datos en la industria.

Modelo	Aplicación
Predicción de calidad de productos	Mejora de ratios de conversión de procesos
Predicción de carga	Optimización de operaciones
Predicción de Control de Procesos	Controles de acción precalculada para sistemas no lineales
Estimaciones en tiempo real	Estimación continua de valores de variables de proceso para definición de parámetros de operación
Estimación de estado de equipos	Optimización del plan de mantenimiento usando estimados de funcionamiento de equipos
Base de conocimiento de operaciones	Acumulación automática del “know how” de operadores expertos

**Tabla 2-3 Otras aplicaciones de minería de datos en la industria**

En el Apéndice se presentan casos documentados de la aplicación de minería de datos a diferentes procesos industriales realizados por la empresa DAEDALUS.

### 3 Desarrollo del Tema

El propósito de este capítulo es presentar el Modelamiento Topológico basado en casos (TCBM), un modelo original basado en Minería de Datos, desarrollado por la corporación japonesa Yamatake en 1993 como apoyo al control avanzado de procesos industriales complejos. El TCBM muestra como los datos históricos acumulados en planta pueden ser usados para desarrollar modelos de estimación o predicción. También se realiza la comparación entre este modelo y los modelos basados en Redes Neuronales. Finalmente, se presentan campos de aplicación donde se ha probado el modelo.

#### 3.1 Modelamiento Topológico basado en casos (TCBM)

El modelamiento topológico basado en casos es una metodología de modelamiento de Caja Negra que emplea tecnologías topológicas (de mapeo) para convertir observaciones históricas o casos en modelos de estimación y predicción basados en minería de datos. El resultado es de alta precisión e índice de confianza.

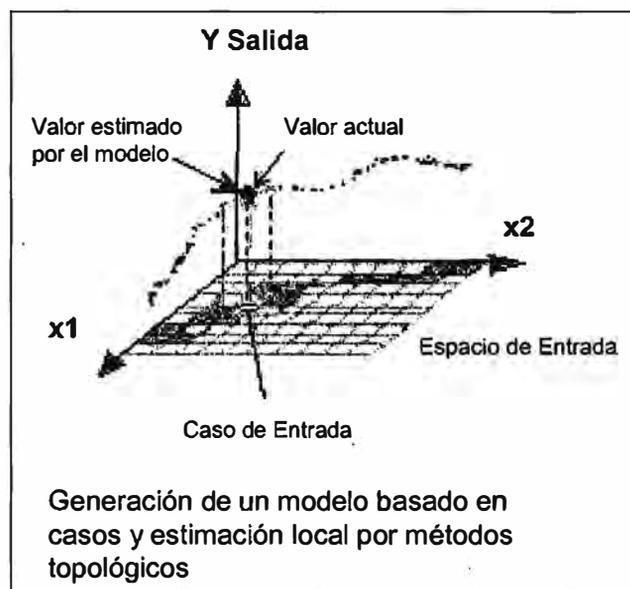


Figura 3-1 Modelo TCBM simple

Por ejemplo, en la figura 3-1 se muestra el método para un modelo simple con solamente dos entradas ( $X_1$  y  $X_2$ ). El modelo trabaja encontrando la salida ( $Y$ ) que empareja lo más cerca posible a un caso de entrada definido por  $X_1$  y  $X_2$ . En general, el método de TCBM utiliza datos históricos para construir un mapa superficial, cada punto del mapa es determinado por un vector dimensional  $X$  de la entrada de  $N$  variables y un valor escalar de salida  $Y$ .

### 3.1.1 El problema típico a resolver

En la figura 3-2 se resume el problema típico de las empresas: se tienen datos acumulados de series de tiempo en la planta, los cuales normalmente no son analizados, dado que es imposible describir una relación directa entre los valores registrados, empleando métodos normales de análisis.

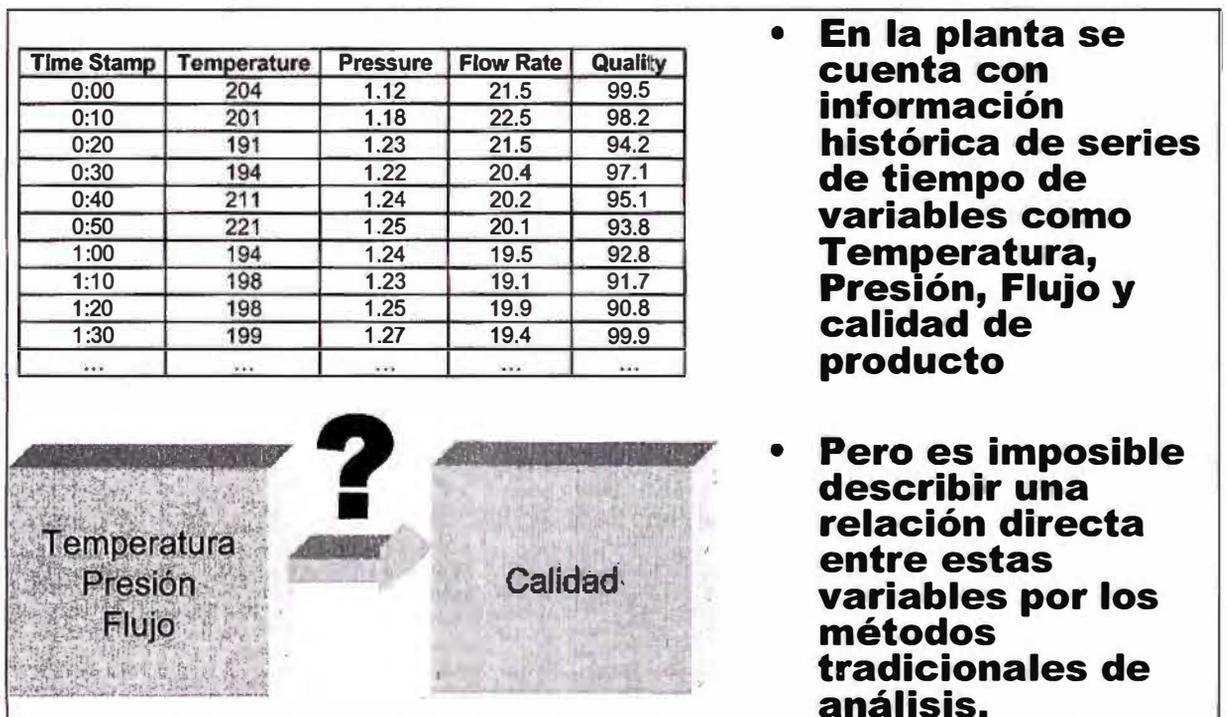


Figura 3-2 Problema típico: Existen datos pero no relaciones

### **3.1.2 La solución: El modelo TCBM**

El TCBM ha sido desarrollado usando conceptos de topología y metodologías de razonamiento basado en casos, y emplea las técnicas de Minería de Datos como soporte informático del modelo.

Usando TCBM se puede construir un modelo usando únicamente datos históricos. Este modelo pertenece a las metodologías de modelamiento denominadas “De Caja Negra”, dado que se basan en conocimientos extraídos y aprendidos de la información analizada, y no en fórmulas o solución de ecuaciones. A este tipo también pertenecen los modelos desarrollados bajo modelos de Redes Neuronales.

Como se describió en la sección 2.2, Minería de Datos, el empleo de esta herramienta permite encontrar patrones y relaciones no evidentes entre los datos analizados; así el modelo recibe la información de planta seleccionada, como por ejemplo series de tiempo de Temperatura, Presión, flujo y calidad, las analiza y encuentra relaciones válidas. Estas relaciones y su grado de validez permiten generar un modelo de estimación o predicción, según sea el algoritmo de minería empleado.

Este nuevo modelo permitirá estimar o predecir valores futuros a partir de datos de entrada.

Como se muestra en la figura 3-3, el TCBM convierte la información disponible en un modelo válido de estimación o predicción:

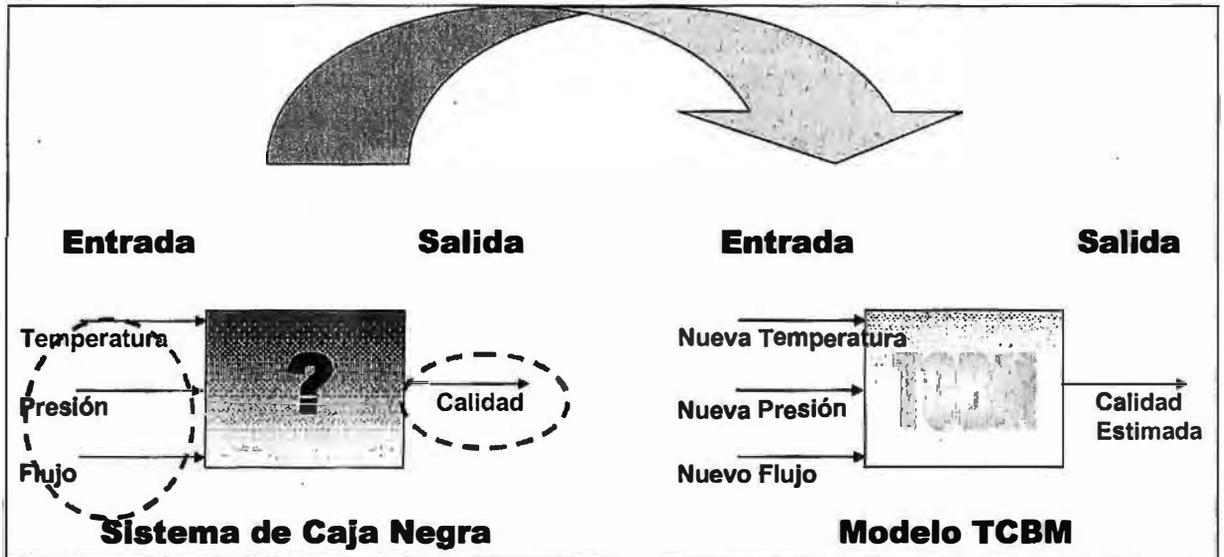


Figura 3-3 Modelo TCBM

### 3.1.3 Características del TCBM

#### *Es una metodología de modelamiento de Caja Negra*

El TCBM es clasificado como una metodología de modelamiento de Caja Negra (solamente las relaciones de entrada-salida son importantes y no como estas son realizadas). Los modelos de aproximación de Caja Negra incluyen a las redes neuronales y sistemas de lógica difusa. Comparado a los modelos mecánicos desarrollados para sistemas complejos, los acercamientos de Caja Negra son a menudo más fáciles de aplicar y el desarrollo del modelo implica un menor esfuerzo. Si para un caso en particular es apropiado emplear un modelamiento de Caja Negra, el emplear TCBM proporciona mejores resultados que los otros modelos de Caja Negra.

#### *Emplea la minería de datos como base del modelo*

El TCBM se basa en un modelo de Minería de Datos. La aplicación de Minería de Datos permite extraer información valiosa, como reglas y relaciones, de los datos almacenados en planta, tal como se muestra en la figura 3-3.

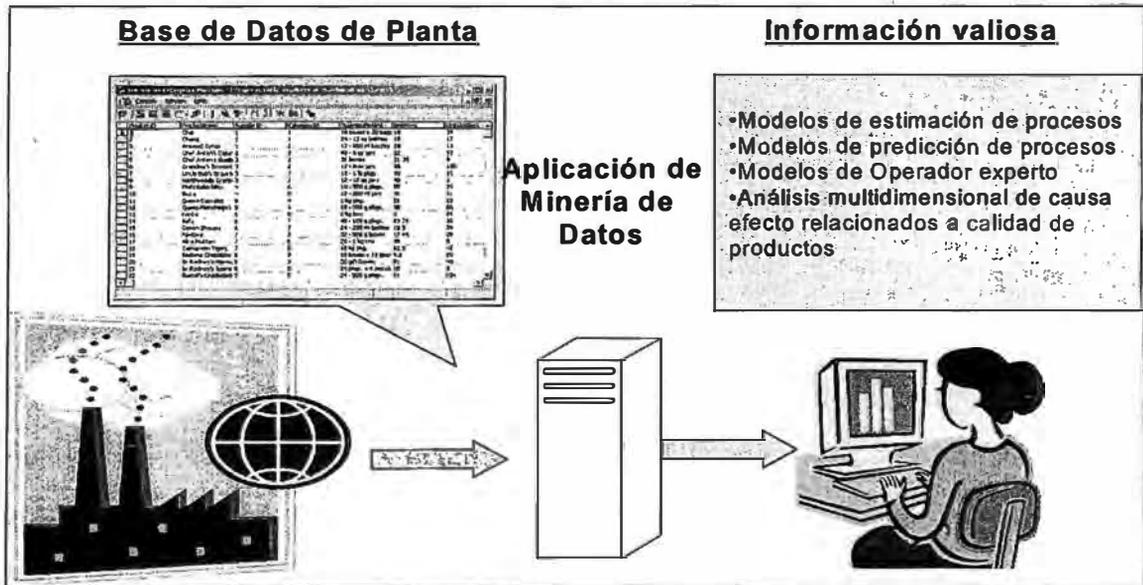


Figura 3-4 Esquema de trabajo del TCBM

### *Índice de Confianza*

El TCBM provee una característica única: el Índice de Confianza. Con el índice de confianza, el usuario sabrá exactamente cuanto puede confiar en el modelo de salida. Esta es una característica crítica para muchas aplicaciones.

El índice de confianza es definido como la distancia al caso más cercano, y es determinado calculando la distancia vectorial entre los valores de entrada actual y el caso más cercano de los datos históricos. Existe un alto grado de confianza si los datos actuales coinciden con un caso previo (distancia = 0). Si el caso actual de los datos de entrada es solamente un vecino cercano a un caso histórico, entonces hay menos confianza en que la salida estimada por el modelo represente la situación verdadera futura.

### *Aprendizaje Adaptativo*

El TCBM también permite una forma simple y conveniente de actualizar el modelo mediante la adición simple de datos a los casos del modelo original

(actualización incremental). Esto resulta en un sistema que aprende y se adapta a las nuevas condiciones y datos del proceso.

### **3.1.4 Proceso de construcción del modelo**

#### ***Preparación para el análisis***

Esta etapa del proceso involucra la extracción de la información de los sistemas donde se encuentra, su depuración y validación. También incluye su almacenamiento en una base estructurada de acorde al algoritmo de minería a emplear.

#### ***Proceso de análisis de variables***

Esta etapa del proceso involucra:

- Realizar cálculos estadísticos básicos
- Análisis de correlación
- Análisis de pasos (delay step)
- Selección de variables lineales
- Selección de variables no lineales
- Selección de variables del modelo
- Validación de patrones y relaciones encontradas

#### ***Modelamiento***

Una vez establecidos los patrones y relaciones entre los datos analizados, se decide si los resultados son válidos para generar un modelo.

Si se determina la validez de los resultados, se decide aplicar un algoritmo de minería y generar un modelo. Construido el modelo, se deben generar datos a partir de variables de entrada, en este punto es importante revisar los datos generados por el modelo y compararlo con la fuente original, para asegurar la

exactitud del modelo. Esta comparación no puede ser hecha al nivel de detalle, pero si a través de muestras comparativas.

A medida que se realicen las consultas al modelo y se valide sus resultados, el modelo irá aprendiendo y otorgando mejores resultados, es decir, con el uso del modelo sucede un proceso de aprendizaje interno, con lo cual en el tiempo se logran mejores resultados.

### ***3.2 Caso práctico: Desarrollo de un modelo para determinar la conversión en la producción de ácido sulfónico***

En esta sección se emplea el software de TCBM para elaborar un modelo que permita estimar la conversión en la producción de ácido sulfónico en una empresa industrial, a partir de la concentración de ácido sulfúrico del producto. El ácido sulfónico es empleado como materia prima para la producción de detergentes (ácido sulfónico lineal) y de pastas dentríficas y lavavajillas (ácido sulfónico ramificado).

El ácido sulfónico es producido por la sulfonación del alquilbenceno, según la siguiente reacción:



El modelo busca estimar el % de conversión de la reacción, es decir el % de materia activa (ácido sulfónico) en el producto final a partir de la medición del contenido de ácido sulfúrico del producto.

Los datos capturados del proceso fueron:

- Tiempo: fecha y hora de la muestra tomada
- Tipo: tipo de producto (L-ácido sulfónico lineal o R-ácido sulfónico ramificado)

- Materia activa: % ácido sulfónico en el producto
- Ácido Sulfúrico: % ácido sulfúrico en el producto

Para el desarrollo del modelo se obtuvieron 4,771 observaciones en un formato de archivo MS Excel. En la tabla 3-1 se muestra un conjunto de las observaciones obtenidas.

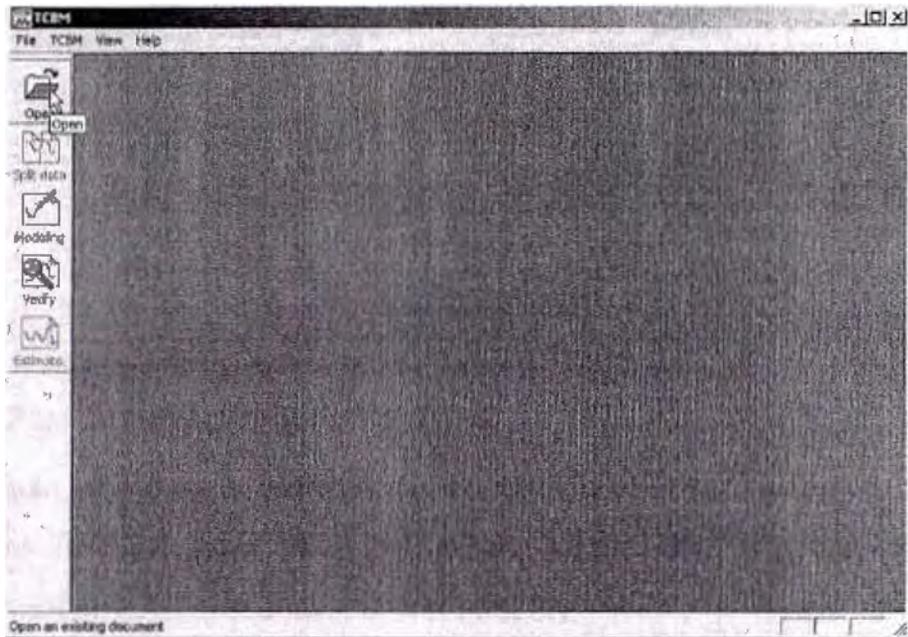
Tiempo	Tipo	Materia Activa	Acido Sulfúrico
7/25/00 10:00	L	96.80	0.92
7/25/2000 13:00	L	96.62	1.27
7/25/2000 16:00	L	97.19	1.20
7/25/2000 22:00	L	96.79	1.07
7/26/2000 4:00	L	97.13	1.06
7/26/2000 10:00	L	96.82	1.11
7/26/2000 13:00	L	96.96	1.08
7/26/2000 16:00	L	97.37	1.02
7/26/2000 22:00	L	97.08	1.14
7/27/2000 4:00	L	96.91	1.06
7/27/2000 10:00	L	96.85	1.14
7/27/2000 13:00	L	97.10	1.16
7/27/2000 16:00	L	96.94	1.09
7/27/2000 22:00	L	96.81	1.06
7/28/2000 4:00	L	97.21	1.12
7/28/2000 10:00	L	97.41	0.96
7/28/2000 13:00	L	97.02	1.14
7/28/2000 16:00	L	97.36	1.09
7/28/2000 22:00	L	97.03	1.19
7/29/2000 4:00	L	97.28	1.14
7/29/2000 10:00	L	96.93	1.08
7/29/2000 13:00	L	97.11	1.15
7/29/2000 16:00	L	97.07	1.10

**Tabla 3-1 Observaciones obtenidas del proceso**

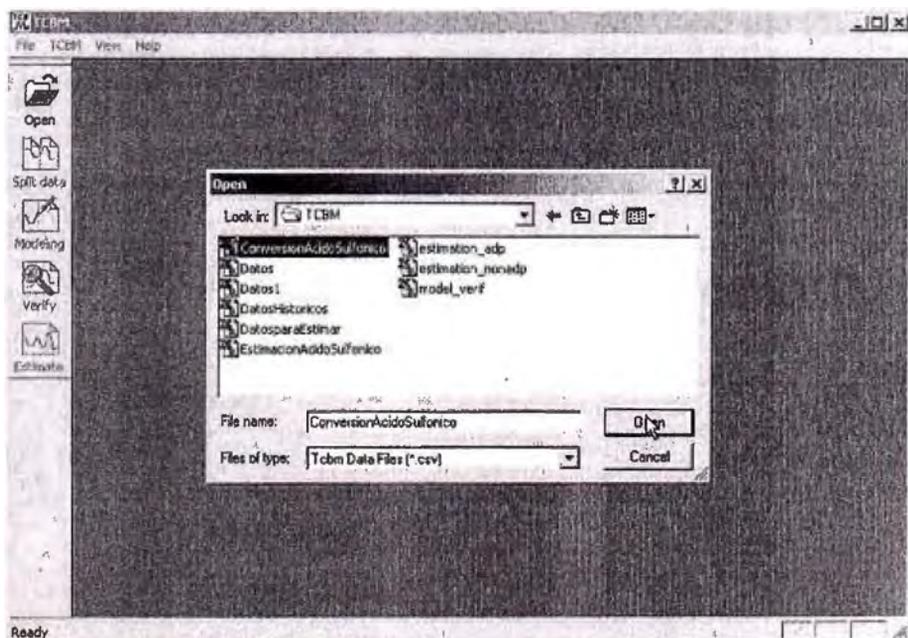
Estos datos son obtenidos del registro diario que hace la empresa en el sistema de control estadístico, luego de realizar las mediciones diarias en los siguientes horarios: 4:00 am, 10 am, 1 pm, 4 pm y 10 pm. Las unidades de medida son % en peso de composición del componente en el producto final.

### 3.2.1 Paso 1: Cargar los datos al software

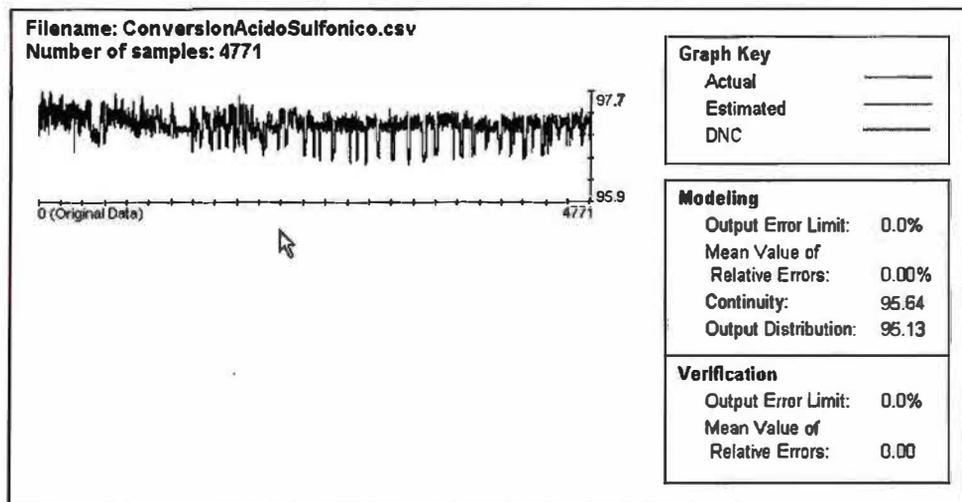
Una vez en el software de TCBM, se procede a cargar los datos históricos mediante la opción “Open”:



Esta opción permite seleccionar el archivo de datos históricos y cargarlos a la herramienta de modelamiento:



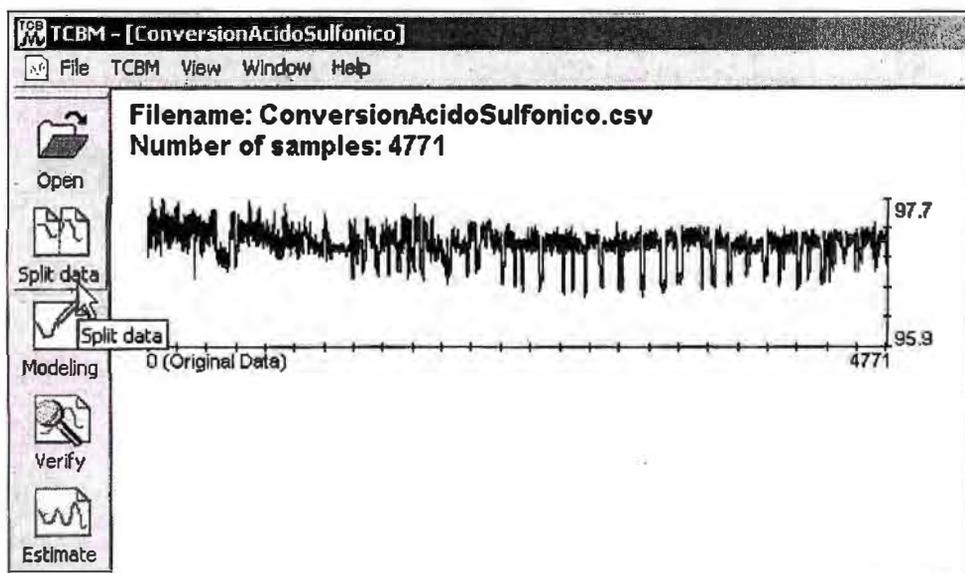
Una vez cargados los datos históricos, estos se presentan en forma gráfica en la herramienta:



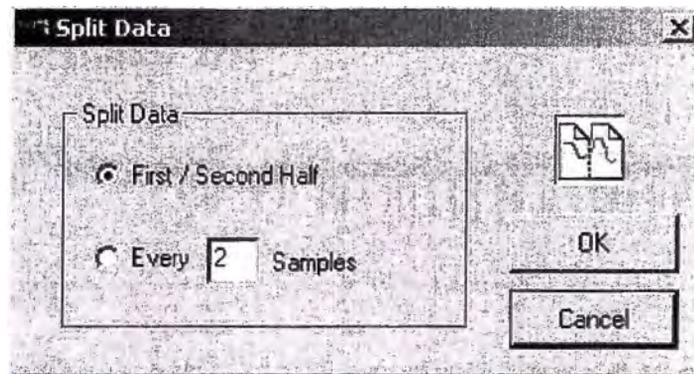
### 3.2.2 Paso 2: Separación de los datos

El segundo paso para la construcción del modelo es la separación de los datos históricos. Esta división se realiza para separar los datos en dos grupos: el primero para construir el modelo y el segundo para realizar la verificación del modelo resultante, esta verificación permite determinar la exactitud del modelo.

Para realizar la separación de datos, se emplea la opción "Split Data" de la herramienta de modelamiento:

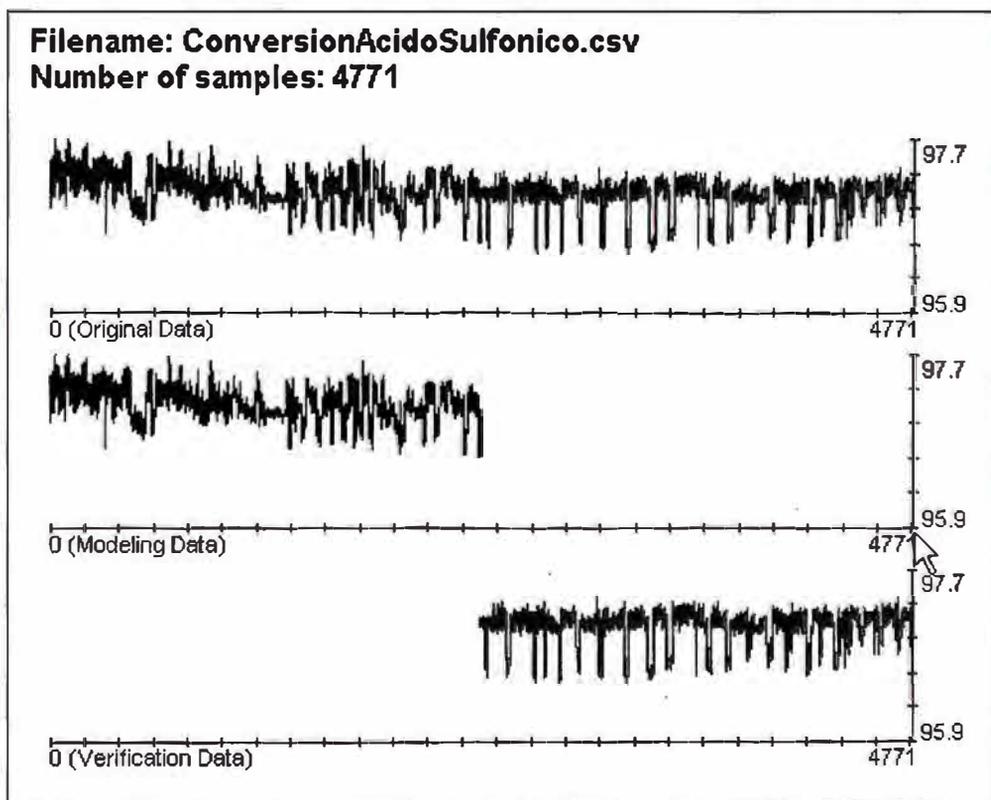


La herramienta permite dos opciones de separación:

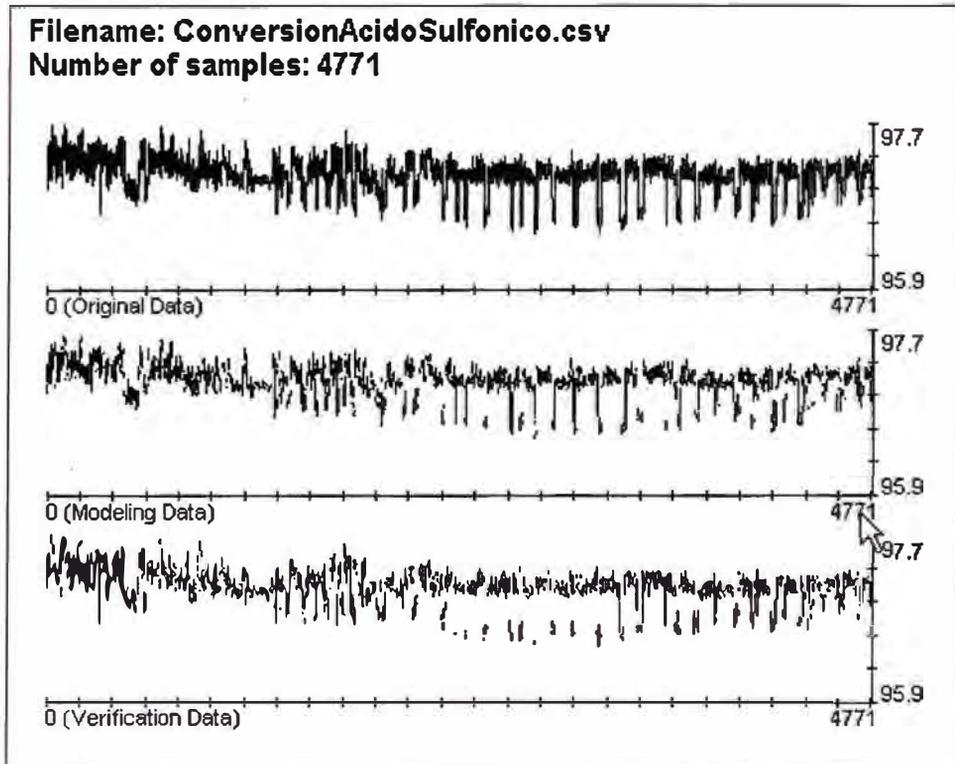


La selección de la división depende de las características de los datos históricos. Para nuestro caso, dado que son muestras de varios meses, y dado que puede haber efectos de serie de tiempos en los datos, seleccionaremos la división cada 2 muestras.

- Primera y segunda mitad, divide los datos en dos secciones iguales



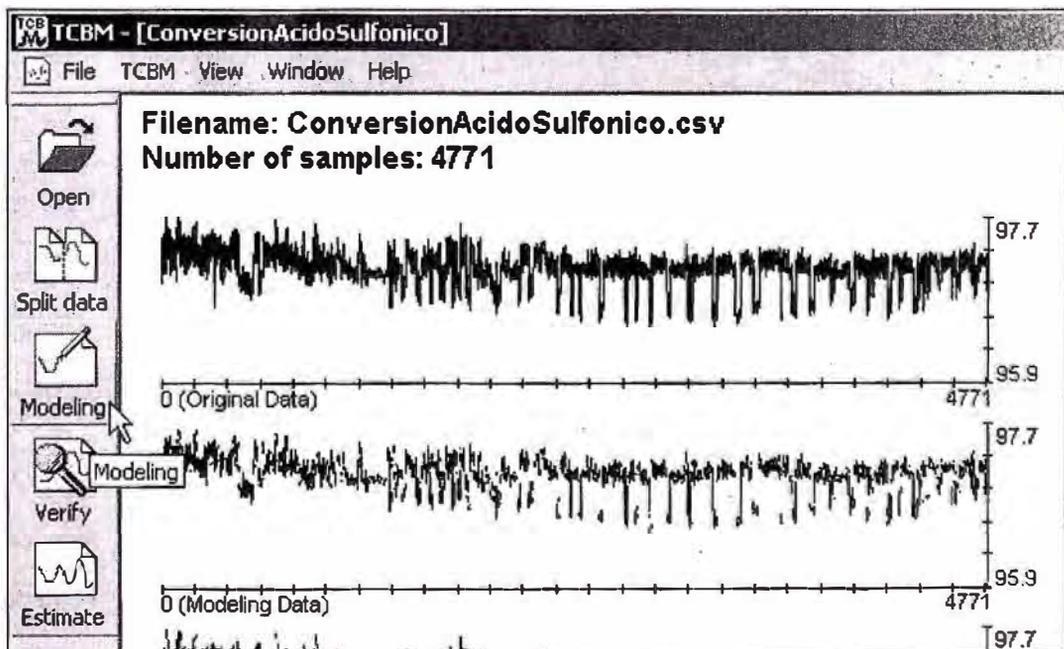
➤ Cada “n” muestras



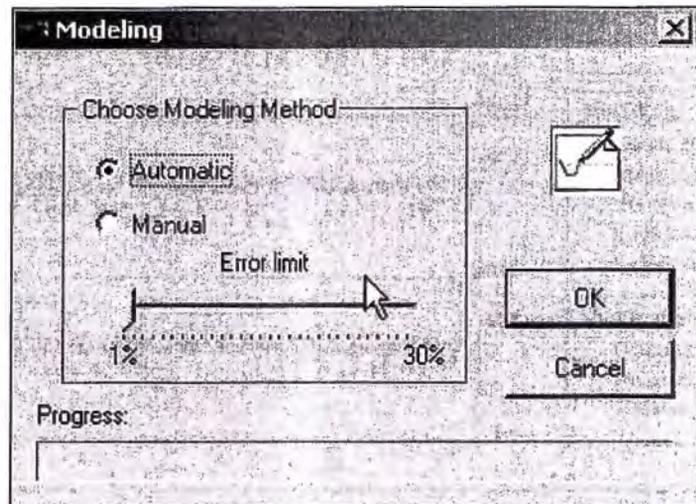
### 3.2.3 Paso 3: Modelamiento

Este paso permite generar un modelo basado en el análisis de los datos históricos.

Para generar el modelo se debe emplear la opción “Modeling”:

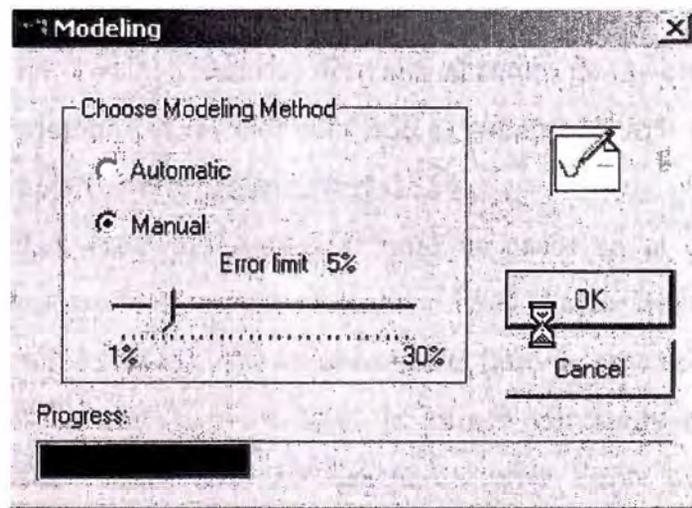


Para generar el modelo, se debe indicar el método a aplicar. Las opciones disponibles son:

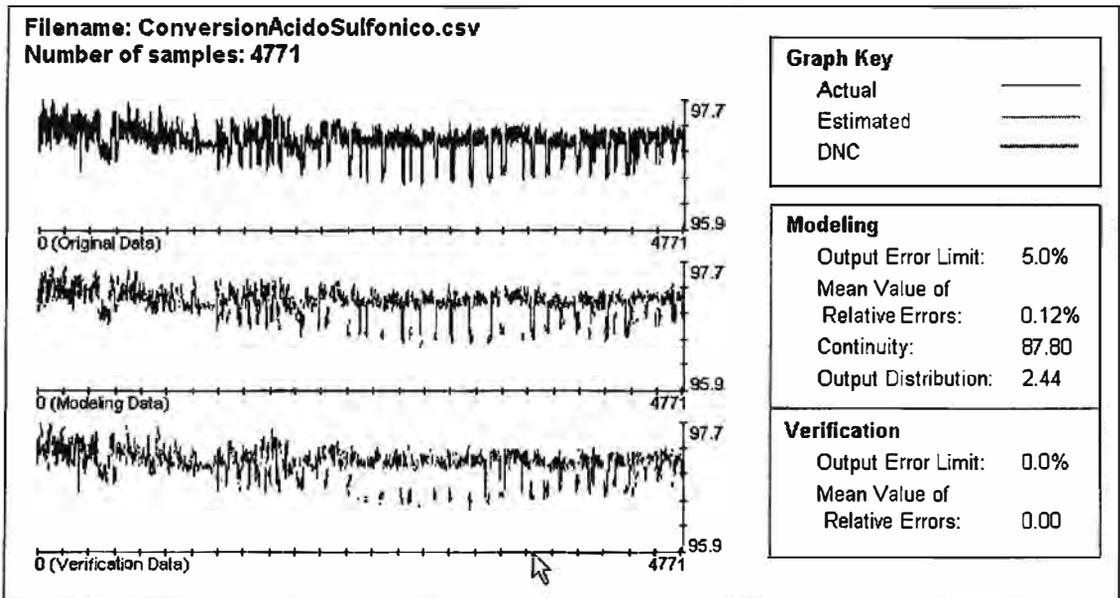


- Automática, en este caso el sistema determina el límite de error
- Manual, en este caso el usuario determina el límite de error permitido

Una vez seleccionado el método a aplicar, presionado el botón "OK", la herramienta inicia la construcción del modelo:



Una vez terminado el proceso de construcción del modelo, se presentan los resultados del modelamiento.



Para verificar la validez del modelo se deben analizar los resultados del modelamiento:

- Limite de error de salida, este es el valor (en porcentaje) que el algoritmo de TCBM usa para definir el modelo. El valor es definido por el usuario o determinado automáticamente por el sistema.
- Valor medio de error relativo, esta es la medida de la exactitud del modelo. Es el valor promedio del valor absoluto de  $(1 - \text{estimado/actual}) \times 100$ . Por ejemplo, si el valor estimado es siempre el 95% del valor actual, el valor medio de error relativo será de 5%.
- Continuidad, este valor indica el ratio de casos en el cual el error de estimación usando el modelo es menor o igual al error límite de salida.
- Distribución de Salida, este valor indica el ratio de casos cuya distribución de valores de salida para los datos de entrada correspondientes pertenecen a los mismos casos que caen dentro del error límite especificado para todos los casos. Si este ratio es alto y el ratio de continuidad es alto también, cada caso cae dentro del error límite de salida y la relación entrada-salida en la vecindad de cada caso es continua.

La tabla 3-2 resume los resultados del modelo generado:

Resultado	Valor obtenido
Límite de error de salida	5.0%
Valor medio de error relativo	0.12%
Continuidad	87.80
Distribución de salida	2.44

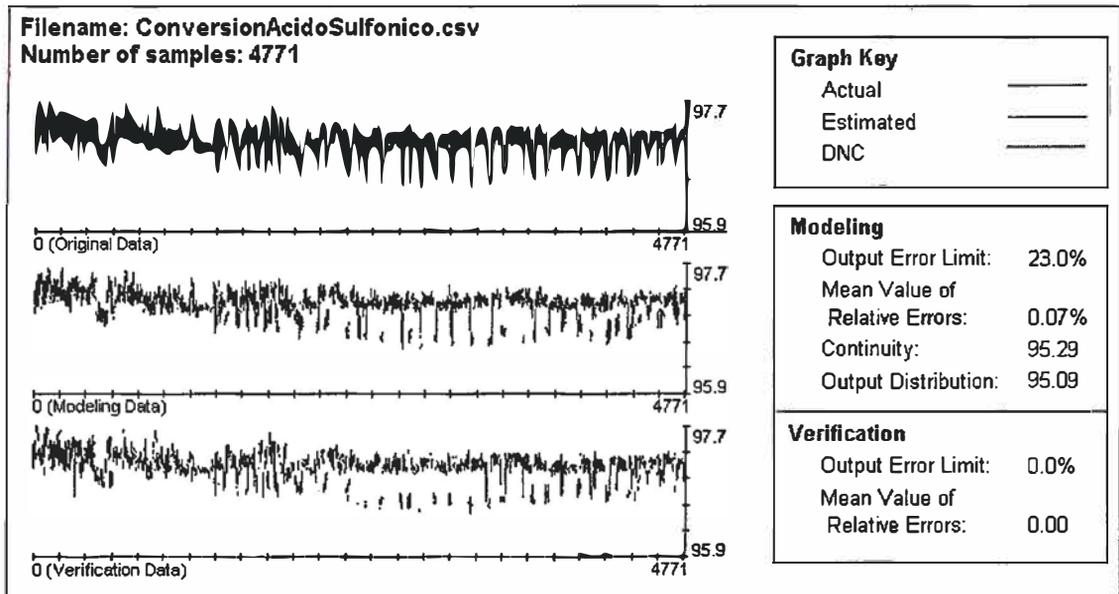
**Tabla 3-2 Resultados del modelo manual (5% error)**

Analizando estos resultados, observamos que si bien el límite de error y valor medio de error relativo son bajos y la continuidad es alta, el valor de distribución de salida es bajo. Esto significa que el modelo tendrá una precisión alta (alto índice de confianza) para predecir valores si los valores de entrada coinciden con algún valor de los datos históricos, pero baja precisión si el valor de entrada esta en la vecindad cercada de un caso histórico.

Para mejorar el índice de confianza del modelo, cambiamos el método de construcción del modelo a automático, para que el pueda encontrar el mejor modelo estimado en función a un límite de error variable. Los resultados obtenidos se presentan en la tabla 3-3:

Resultado	Valor obtenido
Límite de error de salida	23.0%
Valor medio de error relativo	0.07%
Continuidad	95.29
Distribución de salida	95.09

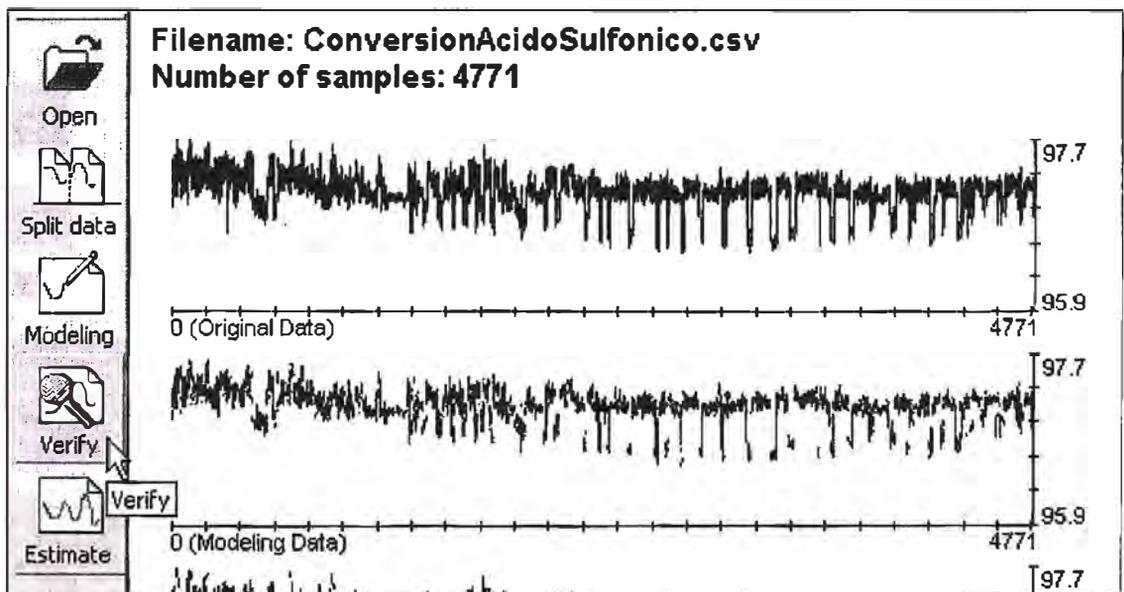
**Tabla 3-3 Resultados del modelo automático**



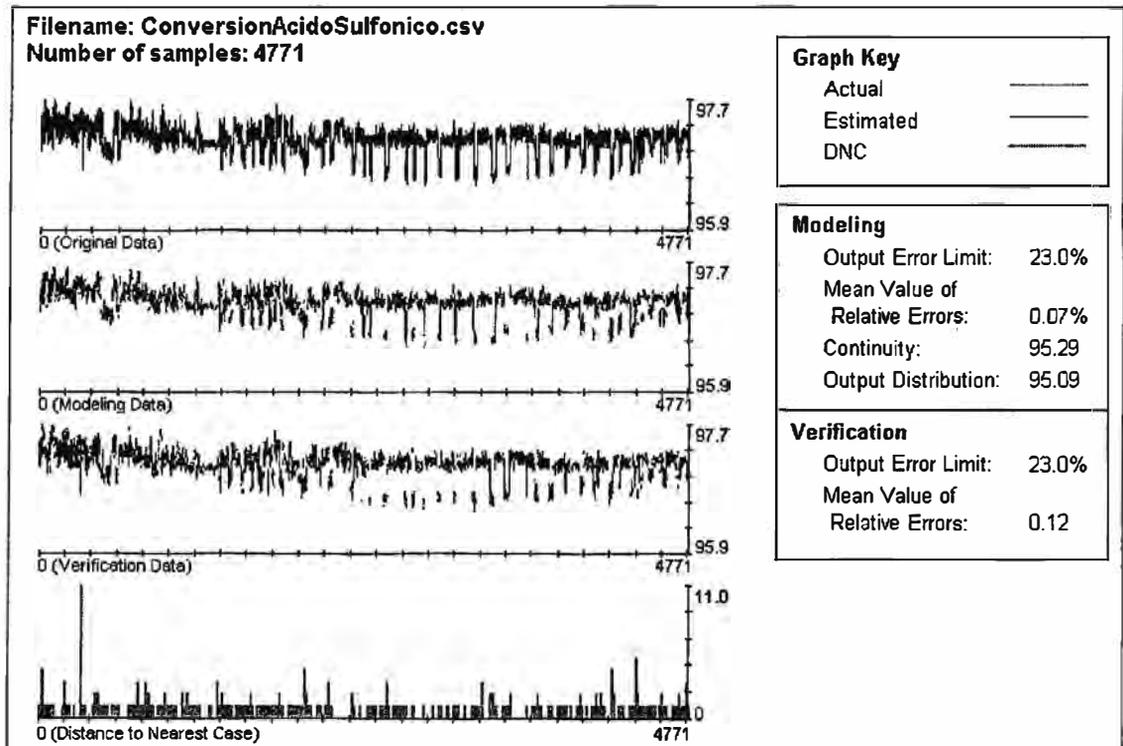
Este nuevo modelo permite obtener un mejor índice de confianza para estimación de cualquier dato nuevo de entrada.

### 3.2.4 Paso 4: Verificación

Este paso permite validar el modelo generado con los datos separados en el paso 2. Para verificar el modelo se debe emplear la opción “Verify”:



Luego, la herramienta presenta los resultados de la verificación del modelo:

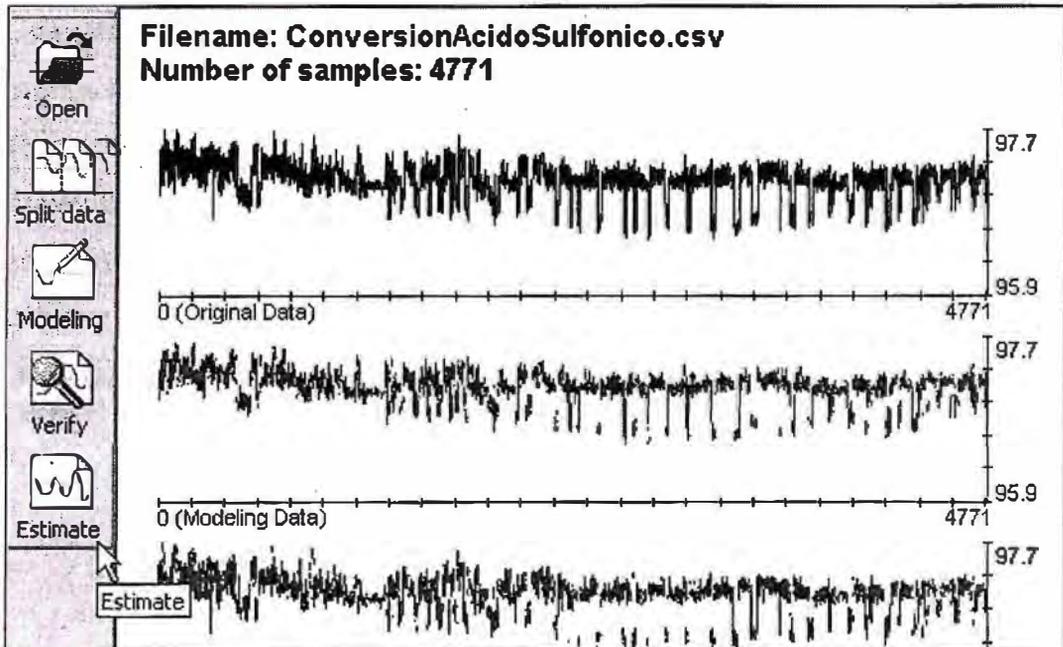


Como se puede apreciar, la verificación ratifica la exactitud del modelo. Se observa que la media de errores relativos es de 0.12% y que el límite de error sigue siendo 23%.

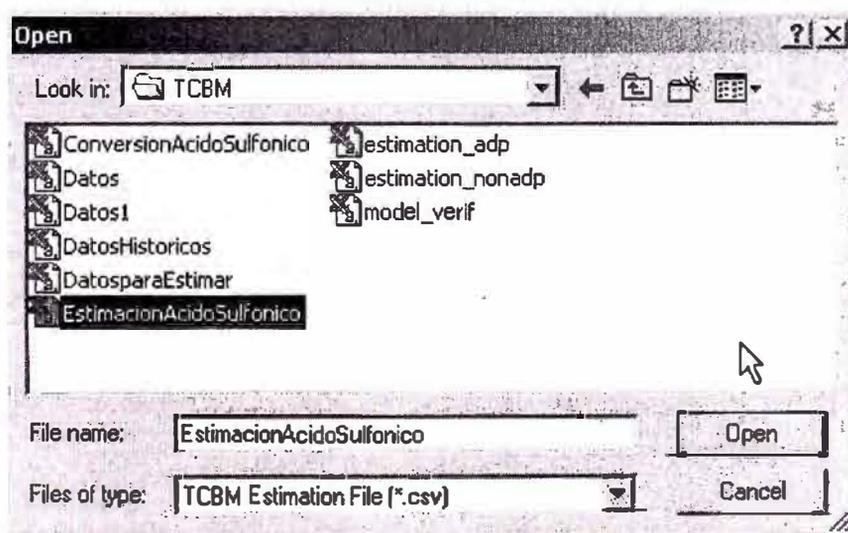
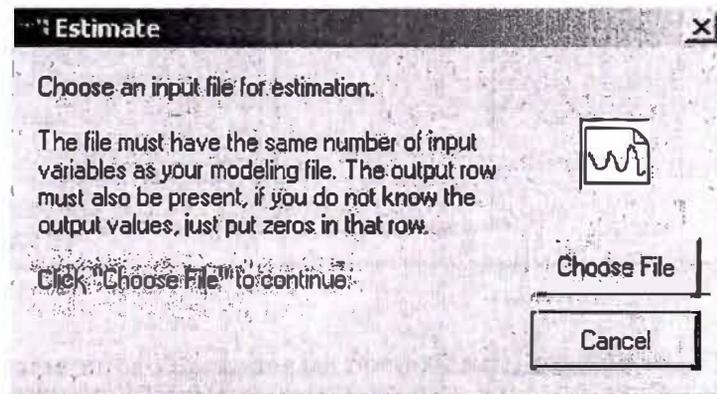
Luego de verificar el modelo y su grado de precisión, ya se puede emplear el modelo para estimar nuevos valores.

### 3.2.5 Paso 5: Estimación

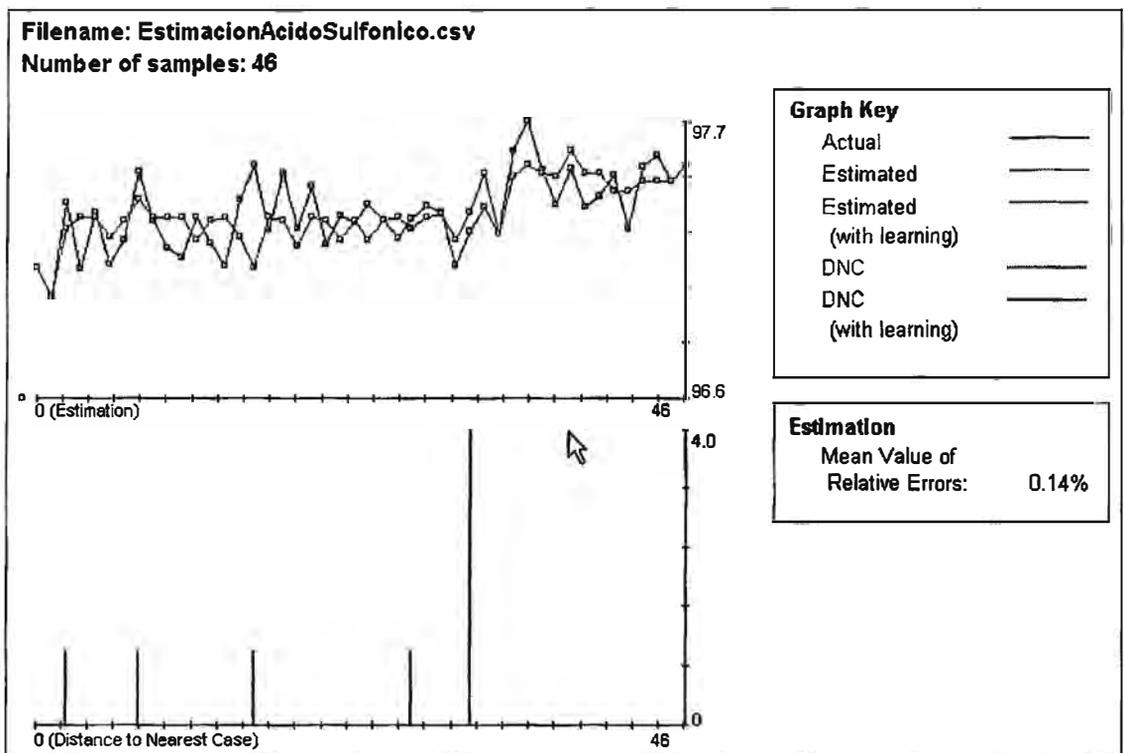
Este paso permite usar el modelo construido para estimar nuevos valores. Para realizar estimaciones se debe emplear la opción “Estimate”:



El modelo permite estimar nuevos valores a partir de un archivo de datos de entrada.



Una vez seleccionado y cargado el archivo con los datos a estimar, el modelo presenta los valores estimados en forma gráfica y tabular. Para este caso práctico se procedió a estimar 46 nuevos valores de conversión de ácido sulfónico a partir de los datos de ácido sulfúrico presente en el resultado del proceso. También se comparó con los resultados reales de ácido sulfónico.



Esta estimación permite comprobar las bondades del modelo.

### ***3.3 Comparación entre metodologías de modelamiento de Caja Negra***

En esta sección se comparan dos metodologías de desarrollo de modelos denominadas “Metodologías de Caja Negra”:

1. Modelamiento empleando Redes Neuronales, el cual se basa en la simulación de la estructura del cerebro humano
2. Modelamiento topológico basado en casos, el cual simula el pensamiento humano.

#### **3.3.1 Modelamiento empleando Redes Neuronales Artificiales**

El cerebro se puede considerar como un computador de alta complejidad, no lineal y de proceso paralelo. Tiene la habilidad de organizar sus neuronas para realizar determinadas actividades, como reconocimiento de patrones, percepción y control motor, en forma mucho más rápida que las computadoras de alta velocidad disponibles en la actualidad.

Una red neuronal artificial (artificial neural network) es un sistema modelado replicando el concepto del cerebro humano. El elemento básico del cerebro humano es la neurona, la cual puede conectarse con otras neuronas para programar una señal determinada. El elemento básico de una red neuronal es una celda computacional denominada “unidad de procesamiento” o “neurona”, la interconexión masiva de estas celdas permite mejorar la performance de estos sistemas, comparados con los esquemas computacionales tradicionales.

En su forma más general, se puede definir una red neuronal como una máquina diseñada para modelar la forma en la cual el cerebro ejecuta una tarea particular o

una función de interés. La red es normalmente implementada usando componentes electrónicos o simulados mediante software de computadora.

Una red neuronal es un procesador paralelo, distribuido y masivo, que tiene una propensión natural a almacenar conocimiento experimental y ponerlo a disposición para su uso. La red se asemeja al cerebro en dos aspectos:

- El conocimiento es adquirido por la red a través de un proceso de aprendizaje
- La fuerte interconexión entre neuronas conocida como peso sináptico es usando para almacenar el conocimiento.

El procedimiento usado para ejecutar el proceso de aprendizaje es denominado algoritmo de aprendizaje.

### ***Ventajas del modelamiento con redes neuronales***

El empleo de redes neuronales para el modelamiento de sistemas ofrece las siguientes ventajas:

- No linealidad. una neurona es básicamente un dispositivo No lineal, consecuentemente, una red neuronal, construida como una interconexión de neuronas, es por si misma No Lineal.
- Existe un mapeo directo entre Entrada-Salida. La base de conocimiento de la red asume que para cada entrada debe existir una única salida, según el problema analizado.
- Adaptabilidad. Una red neuronal tiene la capacidad incorporada de adaptar su peso sinóptico a cambios en el ambiente.

- **Respuesta evidencial.** En el contexto de patrones de clasificación, una red neuronal puede ser diseñada para proveer información no solo del patrón seleccionado sino también de la confianza, o grado de confianza, de la decisión realizada. Esta información puede ser empleada luego para rechazar patrones ambiguos, lo cual mejora el proceso de clasificación de la red.
- **Información contextual.** El conocimiento es representado por un muy estructurado estado de activación de la red neuronal. Cada neurona en la red es potencialmente afectada por la actividad global de todas las neuronas de la red. Consecuentemente, la información contextual es tratada con naturalidad por la red.
- **Tolerancia a fallas.** Una red neuronal, implementada en forma estricta, tiene el potencial de ser inherentemente tolerante a fallas, en el sentido que su performanse se degrada paulatinamente ante condiciones adversas de operación.
- **Implementación a gran escala.** Una red neuronal es ideal para implementaciones en equipos de alta tecnología, de gran capacidad integrada, gracias a su naturaleza de proceso masivo en paralelo. Esto la convierte en una herramienta para aplicaciones de tiempo real, tales como reconocimiento de patrones, proceso de señales y control.

### ***Problemas del modelamiento con redes neuronales***

Los problemas que presenta el modelamiento con Redes Neuronales se centran en:

- Selección de la estructura óptima del modelo, es decir, no hay un procedimiento para establecer el tamaño del modelo: ¿Cuántas neuronas o cuántos niveles de propagación debe contener el modelo?
- Cuanta información es necesaria para tener un buen modelo que pueda ser generalizado. Dado que el modelo se basa en el procedimiento de aprendizaje, es difícil determinar cuando el modelo ha aprendido lo suficiente para poder ser usado.
- Determinar el grado de error incluido en los resultados brindados por el modelo, dado que los datos son convertidos en conocimiento, es imposible entender luego el procedimiento por el cual el modelo determina el estimado o valor esperado.

### **3.3.2 Modelamiento topológico basado en casos**

El modelo topológico basado en casos (TCBM) emplea la metodología de razonamiento basado en casos como base del modelo, y conceptos de topología para solucionar problemas de dimensionamiento propios del modelo de casos.

#### ***Metodología de razonamiento basado en casos***

La metodología de razonamiento basado en casos simula el juicio humano, en otras palabras, la manera de pensar del ser humano.

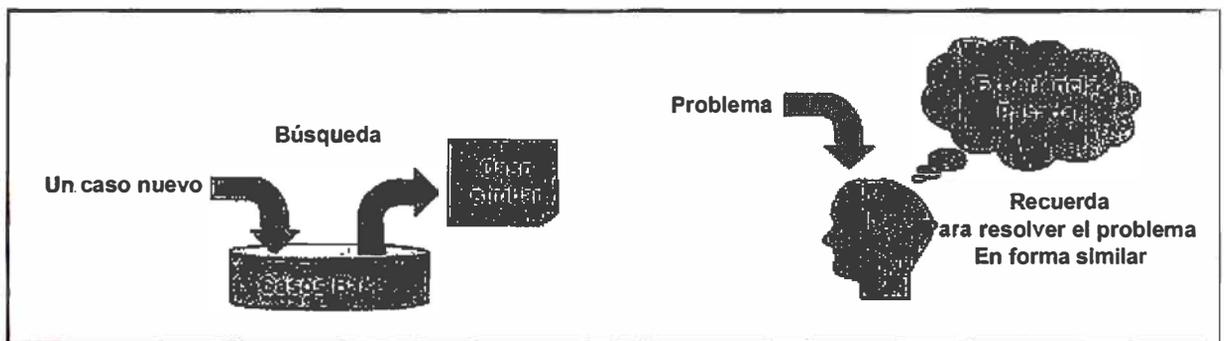
La información analizada es convertida en “casos”, los cuales son almacenados por el modelo como “casos base” o de referencia.

Un caso es un conjunto de datos de entrada-salida. Como lo muestra la figura 3-5, el conjunto de casos representa la “experiencia pasada”.



**Figura 3-5 Razonamiento basado en casos: los casos representan la experiencia pasada**

Cuando se realiza la estimación de valores de salida, se buscan los casos similares (ceranos) y se emplean como base de la estimación. Como se observa en la figura 3-6, el proceso de solución de un problema es similar al razonamiento humano: dado el problema, una persona trata de resolverlo en función a su experiencia pasada.



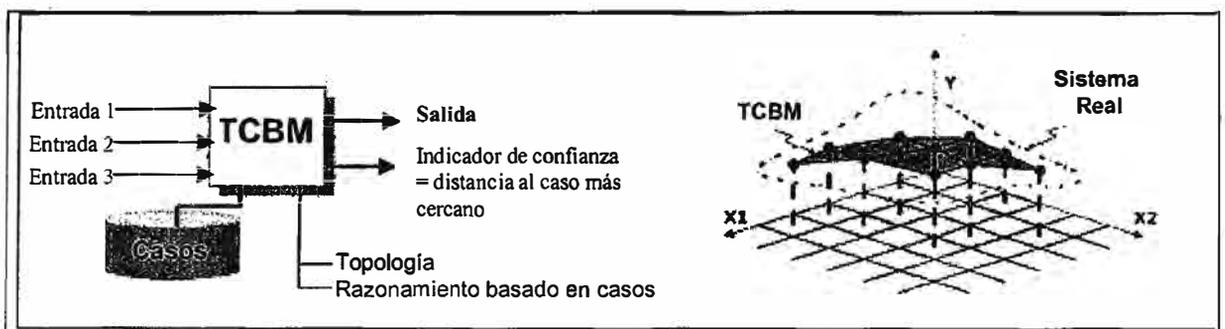
**Figura 3-6 Esquema de resolución de problemas aplicado en el razonamiento basado en casos**

La desventaja de este modelo radica en su dificultad para establecer en que medida existe o no similitud entre los casos o entre los datos en un mismo caso.

Para resolver esta desventaja el TCBM emplea la topología.

### **Topología**

La topología es empleada para realizar la medida de similitud entre los casos (relación). Para establecer esta relación, la topología asume que existe una relación continua entre las entradas y salidas analizadas en cada caso. Como se muestra en la figura 3-7, se emplean algoritmos topológicos para llevar los casos a un plano topológico multidimensional.



**Figura 3-7 Los casos se llevan a un plano topológico dimensional**

El grado de similitud depende de la exactitud del modelo (estimación de error de salida).

### **3.3.3 El modelamiento topológico basado en casos resuelve los problemas del modelamiento basado en redes neuronales.**

Los problemas del modelamiento basado en redes neuronales se reducen a:

1. ¿Cómo determinar la mejor estructura neuronal?
2. ¿Cómo determinar cuantos datos necesito para obtener un buen modelo?
3. ¿Cuánto error esta incluido en los valores estimados de salida?

Estos problemas se eliminan cuando empleamos el modelamiento basado en casos, debido a que:

1. La estructura del modelo basado en casos no requiere una relación funcional entre la entrada y salida, sólo necesita volumen de datos a analizar para establecer patrones y relaciones
2. El modelo basado en casos se basa en el análisis de información previa existente. La validez del modelo se determinará por la precisión de sus resultados en la estimación de valores de salida, no por el número de datos analizados.
3. El grado de confiabilidad del modelo, o por diferencia, el error de estimación, se determina en la validación del modelo con muestras de los datos originales. El proceso de validación determina si el modelo es aceptable o no y su grado de confiabilidad.

### ***3.4 Aplicación práctica de modelos TCBM en la automatización y control de procesos***

#### **3.4.1 Sensor virtual para la medición de concentración de NO<sub>x</sub>**

La medición de la concentración de NO<sub>x</sub> es un proceso difícil de realizar en línea. Por ello, la utilidad de emplear un sensor virtual para determinar esta concentración. El sensor virtual permite estimar los valores de concentración de NO<sub>x</sub> que se obtendrán a la salida del proceso basado en el valor de otras variables, las cuales son de fácil medición.

En la figura 3-8 se diagrama el horno a gas del proceso. Existen medidores para la temperatura del horno y los flujos de ingreso de gas y aire. También se muestra el sensor real de NO<sub>x</sub>, el cual será reemplazado por el sensor virtual.

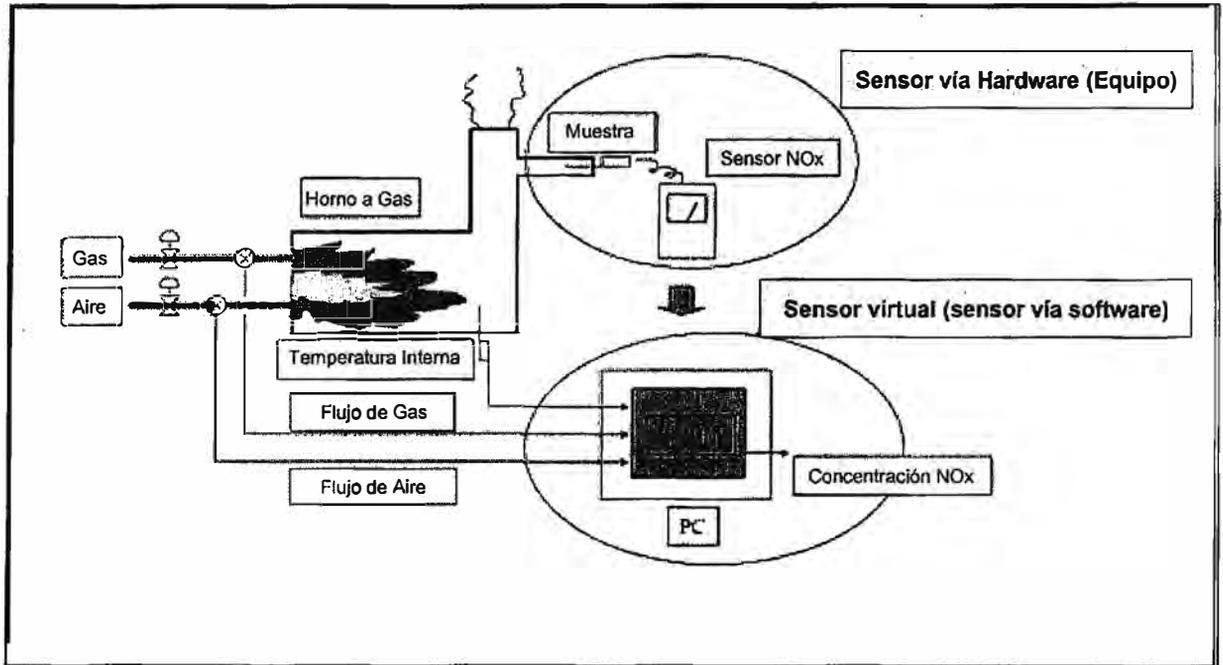


Figura 3-8 Sensor virtual para determinar concentración de NOx

#### Pasos para la construcción del sensor virtual:

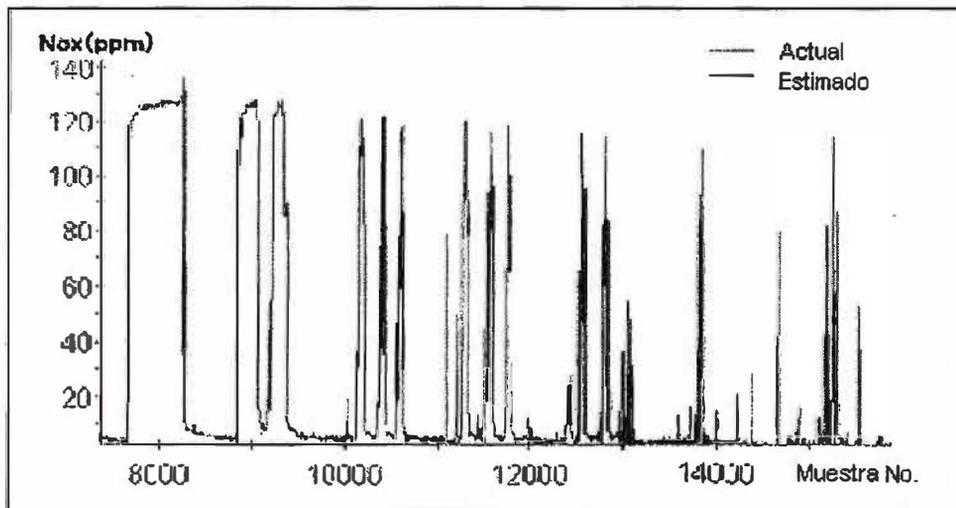
1. Extraer datos históricos de: Temperatura del horno, flujo de ingreso del gas, flujo de ingreso de aire, concentración de NOx resultante (medido por el sensor normal)
2. Construir un modelo de minería de datos basados en la información extraída.
3. Validar el modelo obtenido con una muestra aleatoria de los datos extraídos. Esta validación permite determinar la exactitud del modelo.
4. Construir aplicación para explotar el modelo obtenido (empleando VB)
5. Reemplazar el sensor normal (hardware sensor) por el sensor virtual, construido en el paso 4.

#### Los beneficios del sensor virtual son los siguientes:

1. Bajo costo de mantenimiento
2. De bajo costo de adquisición (o implementación)
3. Permite un monitoreo continuo de la concentración de NOx

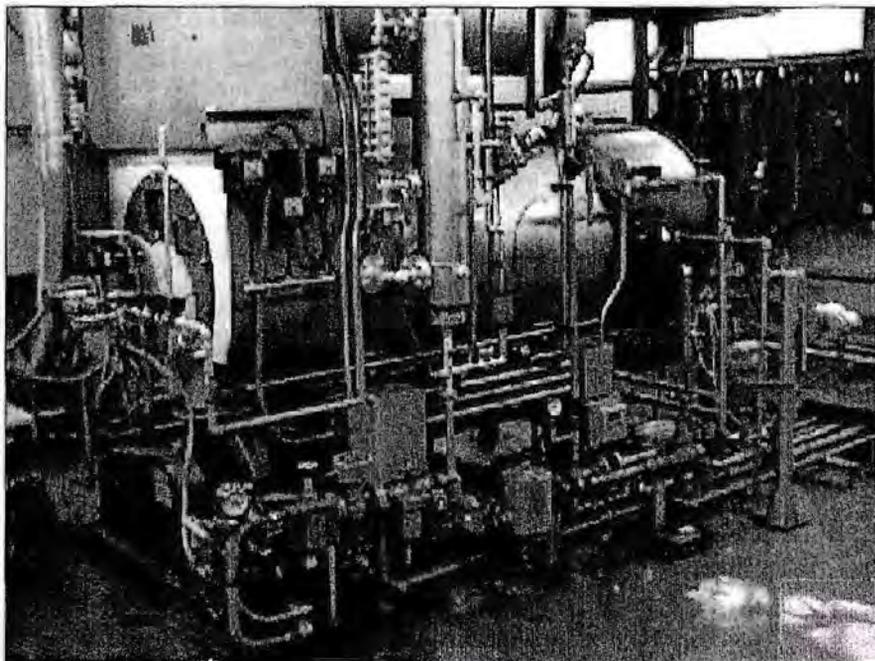
**Resultados del modelo construido:**

Como se observa en la figura 3-9, los resultados del sensor virtual son prácticamente iguales a los valores medidos por el sensor real.



**Figura 3-9 Medición de NOx: Estimado vs. Real**

La figura 3-10 presenta el equipo en planta empleando el sensor virtual:



**Figura 3-10 Horno con sensor virtual de NOx**

### **3.4.2 Determinación de causa de falla para mantenimiento de equipos (empresa Weyerhaeuser)**

Una característica fundamental del software de información de planta instalado en la empresa Weyerhaeuser es la habilidad de capturar y reportar cualquier tipo de evento en el piso de planta. El área de mantenimiento utilizó esta capacidad para registrar las horas de operación de los 102 motores instalados en planta.

Inicialmente la empresa definía los intervalos de mantenimiento en función a las horas de operación estimadas de cada motor. Basado en estos datos se brindaban servicios de mantenimiento a los diferentes centros de trabajo cada tres meses. Sin embargo, aún con este mantenimiento regular, se tenían roturas imprevistas, especialmente en algunos centros de trabajo críticos. Cada falla costaba a la empresa un valioso tiempo de producción.

Por esta razón, se decidió analizar la información de la operación de los motores para determinar las causas de roturas imprevistas.

#### **Pasos para la construcción del modelo:**

1. Extraer datos históricos del sistema de información referidos a la operación de motores, lo cual incluye horas de operación, número de arranques y paros y mantenimientos realizados. La tabla 3-4 presenta una muestra de los datos obtenidos del proceso.

Motor Description	Runtime Hours	Count Of Motor Starts
602200-B6-M1	8.47	4156
602200-B6-M7	22.32	1272
602240-B7-M3	2.73	829
602255-M1	51.07	775
602255-M100	14.27	8979
602255-M2	42.69	52
602255-M3	4.67	5415
602255-M300	21.14	75
602255-M401	107.20	8837
602255-M402	87.07	739
602260-M1	4.31	6268
602260-M102	53.03	6618
602260-M2	3.13	569
602260-M201	51.75	329
602260-M202	82.04	497
602260-M203	81.35	489
602260-M301	37.03	1674
602260-M302	25.32	5923
602260-M401	1.40	512
602260-M402	4.31	2896
602260-M501	85.05	547
602260-M502	86.23	549
602265-M2000	67.83	5854
602265-M2100	4.46	5957
602265-M2200	132.50	5838
602265-M2300	78.17	5778
602265-M2400	40.21	11151
602265-M2500	5.57	8726
602265-M2700	54.60	18051

Tabla 3-4 Datos de operación de motores

2. Construir un modelo de minería de datos basados en la información extraída. El modelo empleado fue el de árbol de decisión.
3. Validar el modelo obtenido mediante la navegación de los nodos del árbol de decisión. Al realizar la validación, el problema se torno evidente, el modelo permitió determinar que la causa de fallas era el número de arranques de cada motor. Esto fue comprobado por el supervisor de mantenimiento al determinar que el número de arranques que originan fallas estaban por encima de las especificaciones técnicas de cada motor.

#### Los beneficios obtenidos por el modelo de determinación de fallas:

1. El modelo permitió determinar que el problema de roturas imprevistas no estaba relacionada con las horas de operación sino con el número de arranques.

2. La empresa ahora emplea los datos de número de arranques y los importa directamente al sistema computarizado de mantenimiento (CMMS) para generar órdenes de mantenimiento basadas en datos concretos y no intuición.
3. Cambio por completo la forma en que la empresa realizaba mantenimiento preventivo y desde que se realizó el cambio, no se han tenido fallas de producción debido a motores. Sólo este cambio ha incrementado la producción de una manera significativa.

### **3.4.3 Determinación de parámetros óptimos de operación para mejorar la calidad de productos**

La empresa SpaceKraft sabía que la calidad de sus productos era afectada por muchos factores, incluidos fallas de equipo, mala calidad de materia primas (papel y pegamento) y condiciones de proceso inestables.

Un estudio inicial permitió determinar que los factores más críticos de la calidad eran la presión a la que se adherían las piezas de cartón y la viscosidad del pegamento. Este estudio no pudo determinar una relación directa entre estas variables.

Por esta razón se decidió analizar la información de operación (presión, viscosidad y calidad de producto) para determinar los parámetros óptimos de operación.

#### **Pasos para la construcción del modelo:**

1. Extraer datos históricos de: Presión, temperatura y Pieza en proceso (Wo Number). Luego se deben completar los datos de la calidad de cada pieza. La tabla 3-5 presenta una muestra de los datos obtenidos del proceso.

Wo Number	Date Stamp	Avg LowPSI	AvgHighPSI	Viscosity Reading	High PressureMax	High PressureMin	Low PressureMax	Low PressureMin
(all)	(all)	(all)	(all)	(all)	(all)	(all)	(all)	(all)
3,294	7/3/02 7:38:55 AM	46.20	52.81	4,373.07	77.07	77.07	40.69	30.26
3,294	7/3/02 7:37:53 AM	46.29	52.47	4,413.33	77.07	77.07	40.35	30.26
3,294	7/3/02 7:38:53 AM	46.20	51.47	4,431.22	77.07	77.07	40.35	30.26
3,294	7/3/02 7:39:52 AM	45.94	50.84	4,319.07	77.07	77.07	39.76	27.92
3,294	7/3/02 7:40:51 AM	45.94	51.20	4,366.61	80.59	80.59	38.04	27.92
3,294	7/3/02 7:41:49 AM	45.94	50.69	4,410.34	80.59	80.59	38.04	27.92
3,294	7/3/02 7:42:49 AM	45.94	50.63	4,427.90	80.59	80.59	38.04	27.92
3,294	7/3/02 7:43:46 AM	45.94	50.60	4,428.90	80.59	80.59	37.69	27.08
3,294	7/3/02 7:44:44 AM	45.94	50.40	4,509.73	80.59	80.59	37.69	27.08

**Tabla 3-5 Datos del proceso de pegado**

2. Construir un modelo de minería de datos basados en la información extraída.
3. Validar el modelo obtenido con una muestra aleatoria de los datos extraídos. La validación permitirá establecer la relación entre la calidad de producto y la presión y temperatura

#### **Los beneficios obtenidos por el modelo:**

1. El modelo ha permitido a SpaceKraft establecer la relación entre las dos variables (presión y temperatura) y de estas con la calidad del producto. Esto ha ayudado a determinar la tolerancia de operación de estas variables en el proceso, para no afectar calidad del producto.
2. También ha ayudado a identificar los parámetros correctos de calidad para cumplir con las especificaciones requeridas por cada cliente.

## 4 Conclusiones y Recomendaciones

### 4.1 Conclusiones

- Con la denominada sociedad de la información se está produciendo un fenómeno curioso, día a día se multiplica la cantidad de datos almacenados. Sin embargo, contrariamente a lo que pudiera esperarse, esta explosión de datos no supone un aumento de nuestro conocimiento, puesto que resulta imposible procesarlos con los métodos clásicos. Nos enfrentamos a la paradoja de que, cuantos más datos están disponibles, menos información tenemos.
- La idea clave es que los datos contienen más información oculta de la que es ve a simple vista, por ello tenemos que usar las Tecnologías de Información disponibles para evidenciar el conocimiento y así formar empresas que aprenden.
- La Inteligencia de Negocios y la Minería de Datos están dentro de estas tecnologías relativamente nuevas que surgen de la necesidad de procesar, analizar y aprovechar esta información oculta en los grandes volúmenes de datos generados.
- La Minería de Datos busca, extrae y analiza enormes cantidades de datos para descubrir patrones y derivar su significado. Involucra tanto la descripción de lo sucedido (¿qué pasó? y ¿porqué pasó?), como la predicción de las tendencias futuras (¿qué pasará?).
- El análisis de la tendencia actual del manejo de la información en la industria y del mercado, permite vislumbrar el increíble potencial de la aplicación de herramientas de Minería de Datos. Existen y pueden desarrollarse cuantiosas aplicaciones para la mejora de los procesos industriales, proyectos de diseño y producción, control de calidad, etc.
- Basado en estas nuevas tecnologías, la corporación Yamatake desarrolló el modelamiento topológico basado en casos (TCBM, por sus siglas en ingles

- Topological case based modeling) como herramienta de apoyo al control avanzado de procesos industriales complejos. El TCBM se basa en el empleo de minería de datos para elaboración de modelos de predicción y estimación; es aplicable a sistemas continuos lineales y no lineales; de fácil implementación y alta precisión.
- La aplicación del TCBM en la Ingeniería Química es inmediata, dado que se sustenta en información histórica disponible en planta para el desarrollo de los modelos.
  - La inversión en elaboración de modelos basados en TCBM, o en otras técnicas o herramientas de minería de datos, se sustenta en los beneficios a obtener, como lo demuestran los casos presentados en el informe. Estos proyectos permitirán a la empresa que los aplique ahorrar dinero mejorando la eficacia de sus actividades, o bien, ganar dinero descubriendo nuevas formas de optimización de procesos o de obtención de beneficios.
  - El modelo desarrollado en el presente informe para estimar la conversión de ácido sulfónico, a partir de la concentración de ácido sulfúrico del producto, es un ejemplo práctico de los beneficios de la aplicación de estas nuevas tecnologías. El modelo permite estimar el grado de conversión de la reacción, y por ende la calidad del producto, a partir de la concentración de ácido sulfúrico en el producto final. El modelo tiene un alto índice de confianza y lo que es más importante, permitirá reducir costos, dado que el análisis actual realizado por la empresa que mide concentración de ácido sulfónico y ácido sulfúrico, se podrá reducir sólo al análisis de ácido sulfúrico, y a partir de este estimar el ácido sulfónico. Esta simplificación en el análisis permite reducir horas-hombre de laboratorio y uso de reactivos como principales beneficios.

## **4.2 Recomendaciones**

Si se desea implementar un proyecto de minería de datos en una empresa, se propone seguir las siguientes recomendaciones:

➤ **Para la selección de un proyecto**

Evaluar la información disponible en la empresa y seleccionar una fuente de información a la que pueda aplicarse minería de datos y que permita generar un modelo para optimizar procesos, reducir costos o mejorar rendimientos. Para seguir con el siguiente paso es imprescindible determinar los objetivos a cubrir con el modelo a generar y sustentar el proyecto en función a su retorno de inversión. También es necesario validar que la información disponible sea suficiente para generar el modelo.

➤ **En la planificación del proyecto**

Antes de iniciar la implementación técnica, es necesario involucrar a las personas que tomarán decisiones basadas en el modelo para entender sus requerimientos de información y determinar el alcance funcional del modelo a implementar. Esta información permitirá planificar la solución. Esta planificación incluye el desarrollo de una visión del proceso de toma de decisiones que soportará el modelo, la fuente de datos que estará disponible para el modelo, el modelo de minería de datos a aplicar, y la forma como el modelo presentará la información al usuario final.

➤ **En la formación del equipo de trabajo**

El equipo de trabajo debe estar constituido por consultores de negocio que entiendan las decisiones estratégicas y tácticas que quieren realizar los usuarios con el modelo, y por consultores técnicos que entiendan como construir soluciones de minería de datos.

➤ **En la implementación del proyecto**

El desarrollo de un proyecto de minería de datos requiere servicios de implementación en tres áreas básicas. El equipo de trabajo deberá contar con especialistas en estas áreas o evaluar subcontratar estos servicios a terceros, para:

1. Selección e Instalación de la Plataforma Tecnológica

Involucra la evaluación de las herramientas disponibles en el mercado y la selección de la más adecuada para el modelo a desarrollar.

2. Desarrollo de la solución

Desarrollo del modelo de decisión basado en el problema a resolver, por parte de los consultores de negocio, y su implantación en la plataforma tecnológica seleccionada, por el equipo técnico. Terminado el desarrollo se debe evaluar si la solución cumple con los objetivos definidos en la planificación del proyecto

3. Documentación y Entrenamiento

Implica la documentación del modelo desarrollado y su forma de uso. La documentación es útil para el posterior mantenimiento de la solución. También se debe capacitar en el uso de la solución a todos los usuarios que emplearán el modelo desarrollado.

➤ **En el desarrollo de nuevos proyectos**

Los resultados obtenidos por el proyecto deben ser usados para motivar a la gerencia en seguir invirtiendo en proyectos de este tipo. En la medida en la que la inversión en inteligencia de negocios se incrementa, lo hace también el valor de los datos de la empresa

## 5 Bibliografía

Seidman C., “Data Mining with Microsoft SQL Server 2000 Technical Reference”, Microsoft Press, Estados Unidos, 2001, Pag. 3.

Merril W., 2003, “Minería de Datos y Modelamiento Topológico basado en Casos (TCBM)”, 1 disco compacto: TECSUP - V Simposio Internacional de Automatización: “Estrategias para la Gestión Eficaz de los Recursos Productivos”

Pulido A., 2003, “Integración de Sistemas de Planta con MES”, 1 disco compacto: TECSUP - V Simposio Internacional de Automatización: “Estrategias para la Gestión Eficaz de los Recursos Productivos”

López M., 2003, “La información como un Elemento de Impulso en la Manufactura”, 1 disco compacto: TECSUP - V Simposio Internacional de Automatización: “Estrategias para la Gestión Eficaz de los Recursos Productivos”

Corporación Yamatake, “Topological Cased Based Modeling (TCBM) Work-Bench”, página web: [www.tcbm-solutions.com](http://www.tcbm-solutions.com).

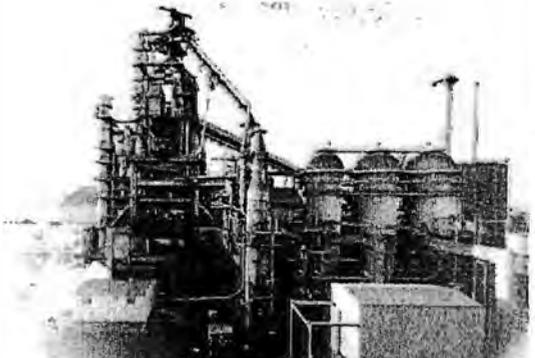
Daedalus - Data, Decisions and Language, S.A., página web: [www.daedalus.es](http://www.daedalus.es).

Mavrovouniotis M., “Artificial Intelligence in Process Engineering”, ACADEMIC PRESS INC., Estados Unidos, 1990, Pag. 189.

Haykin S., “Neural Network, a Comprehensive Foundations”, Hamilton Printing Company, Estados Unidos, 1994, Pag. 1.

## **6 Apéndice: Casos documentados de aplicación de minería de datos a procesos industriales**

- Minería de datos para la optimización de altos hornos
- Minería de datos en optimización de la Producción de Cartón
- Minería de datos para Gestión de Alarmas en Plantas Petroquímicas
- Minería de datos en la Fabricación de Electrodomésticos
- Minería de datos en la Producción de Cemento
- Minería de datos para Control de Calidad de Materiales

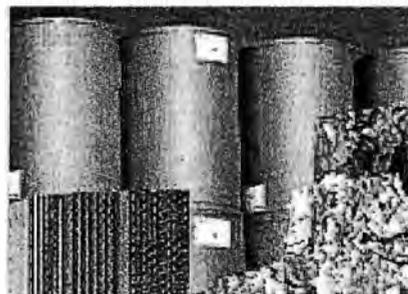
<b>NOMBRE DEL PROYECTO</b>	Aplicación de minería de datos al control y optimización de altos hornos
<b>SECTOR</b>	Industria - Altos hornos
<b>OBJETIVO</b>	Desarrollo de un sistema de optimización/control de procesos complejos. Este sistema se ha aplicado en altos hornos, obteniéndose una reducción apreciable del consumo de combustible.
<b>DESCRIPCIÓN</b>	 <p>Actualmente, los altos hornos incluyen exhaustivos sistemas de medida que permiten supervisar y controlar su funcionamiento. Los datos sobre la temperatura y la presión en el horno, así como los resultantes del análisis de muestras de la salida de gases, se utilizan en muchos hornos modernos para optimizar el proceso de fundición. Sin embargo, las relaciones existentes entre dichos datos no pueden describirse mediante modelos analíticos, por lo que se necesita un operador experto que analice e interprete las medidas.</p> <p>El sistema desarrollado permite automatizar estas tareas mediante la aplicación de técnicas de minería de datos. Se basa en el análisis de la distribución de temperaturas en la parte superior del horno, que, junto con otras medidas complementarias, proporciona indicadores característicos de la calidad del proceso. Aplicando métodos híbridos neuro-borrosos, se ha construido un modelo de la interdependencia entre los parámetros de operación del proceso y los perfiles de temperatura correspondientes.</p> <p>El sistema funciona integrado en la línea de producción. Utilizando técnicas de <i>clustering</i> (agrupamiento) borroso, analiza los datos recopilados para determinar el estado en que se encuentra el horno. Dicho estado se compara con el que se pretende conseguir, analizándose la desviación mediante una red neuronal. Basándose en los datos recibidos, la red neuronal formula una recomendación sobre las correcciones a realizar en la relación carga-combustible empleada en el sistema de carga. El operador evalúa esta recomendación y la transforma en medidas concretas, seleccionando para ello uno de los programas de carga definidos mediante el sistema borroso.</p> <p>En una etapa previa se determinaron los distintos modelos a utilizar según la situación. Basándose en los archivos históricos de datos, se definieron las diferentes estrategias de solución que se aplicarían en cada caso. Estos modelos son los que utilizará el sistema inicialmente, pero pueden ser optimizados en función de su comportamiento durante el funcionamiento operativo del sistema.</p> <p>Este sistema, denominado ROSA (Computer Controlled Measuring Analysis), se halla en funcionamiento en dos altos hornos y ha supuesto una reducción del consumo de combustible de aproximadamente 1 kg. de coque por tonelada de lingotes de hierro, con un ahorro resultante de unos 380.000 euros al año.</p> <p><i>Este sistema ha sido desarrollado por Dango &amp; Dienenthal Maschinenbau GmbH utilizando la herramienta DataEngine. DataEngine es distribuido en España por DAEDALUS – Data, Decisions and Language, S.A. Nuestro departamento de Ingeniería está a su disposición para desarrollar proyectos como éste.</i></p>

# Minería de Datos en Optimización de la Producción de Cartón

<b>NOMBRE DEL PROYECTO</b>	Aplicación de minería de datos para la optimización del proceso de producción de cartón
<b>SECTOR</b>	Industria – Papelera
<b>OBJETIVO</b>	Mejorar el rendimiento de la fabricación de cartón optimizando el control de la velocidad de la línea.

## DESCRIPCIÓN

Para la fabricación de cartón, la pasta inicial preparada a partir de papel desechado se somete a una serie de tratamientos, tanto mecánicos como químicos, a lo largo de una línea de proceso. La suspensión de fibra inicial es filtrada para eliminar una parte del agua y obtener la malla del papel. Dicha retícula se somete a un prensado y, posteriormente, a un proceso de evaporación, para eliminar el resto de agua que aún contiene, utilizando una serie de cilindros calentados mediante vapor.



En cada una de las secciones pueden ajustarse una serie de parámetros que afectan tanto a la calidad del producto como al rendimiento del proceso. Este tipo de líneas utilizan un control de proceso que es puesto a punto inicialmente por ingenieros y sometido a un ajuste más exacto por los operadores de línea. Sin embargo, con este planteamiento los ajustes dependen en gran medida de la experiencia y opinión personales del técnico que los realice.

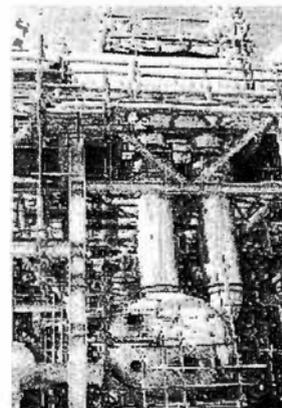
En este proyecto se han aplicado técnicas de minería de datos, en concreto de *clustering* (agrupamiento) borroso, para obtener un modelo del proceso que aclarara las relaciones entre los diferentes parámetros y permite extraer una serie de pautas y unificar los criterios aplicados por los operadores. Las conclusiones cuantifican una serie de dependencias, muchas ya previstas pero que no habían podido ser comprobadas experimentalmente debido al riesgo de pérdida de calidad y fallos en la producción, y otras desconocidas hasta ese momento.

El análisis ha permitido identificar cinco estados, correspondientes a otros tantos modos de funcionamiento de la línea según las características que se consideren prioritarias (calidad, ahorro de vapor, velocidad...). Para cada uno de ellos se han obtenido los valores típicos de los parámetros de operación, que permiten deducir las relaciones entre ellos, en especial con la velocidad de la línea.

El estudio realizado ha permitido mejorar especialmente la velocidad en tres de dichos estados, con lo que se ha conseguido una mejora global del 1%, lo que significa un ahorro de aproximadamente medio millón de euros al año.

*Este sistema ha sido desarrollado por Datenservice Jost utilizando la herramienta DataEngine. DataEngine es distribuido en España por DAEDALUS – Data, Decisions and Language, S.A. Nuestro departamento de Ingeniería está a su disposición para desarrollar proyectos como éste.*

<b>NOMBRE DEL PROYECTO</b>	Aplicación de minería de datos a la gestión de alarmas en plantas petroquímicas
<b>SECTOR</b>	Industria - Petroquímica
<b>OBJETIVO</b>	Desarrollo de un sistema de control para supervisión del estado del horno en procesos de producción de etileno.
<b>DESCRIPCIÓN</b>	<p>Más del 97% del etileno producido anualmente en el mundo se obtiene por craqueo térmico con vapor de hidrocarburos del petróleo. La nafta, obtenida por destilación del petróleo crudo, es la principal materia prima del etileno. Se calienta en hornos de craqueo hasta 820-840°C, temperatura a la que se produce la reacción química.</p> <p>Durante el proceso también se producen acetilénicos, diolefinicos y componentes aromáticos que provocan que se deposite coque en la superficie interna de las tuberías del horno. Esta capa de coque obstaculiza la transferencia de calor al gas, por lo que cada cierto tiempo debe pararse el horno para eliminarla. Para garantizar el funcionamiento ininterrumpido de la planta, suelen utilizarse varios hornos en paralelo en el proceso.</p> <p>El problema es que los parámetros que se miden en el proceso, que recogen el valor de diferentes indicadores como temperaturas, presiones o flujos, no permiten determinar por sí solos el estado del horno. El operador decide cuándo debe realizarse una limpieza basándose en el tiempo de funcionamiento del horno en cuestión y en su propia experiencia. Este método es, obviamente, inexacto, de ahí el interés de conseguir un sistema de control que ayude en la toma de decisiones.</p> <p>Puesto que no se dispone de un modelo determinístico que relacione dichos parámetros, se aplican técnicas de minería de datos para extraer la información útil oculta en ellos, en este caso, si existe o no exceso de coque. Utilizando métodos de <i>clustering</i> (agrupamiento) se puede extraer información de los conjuntos de datos, determinando clases que pueden utilizarse para determinar el estado del proceso; estados similares se asignan a una misma clase. El estado de un proceso se describe mediante un conjunto de características que representan las variables cuantitativas medidas en el sistema, como temperaturas, etc. En este caso, se han utilizado técnicas de <i>clustering</i> borroso para detectar similitudes entre los estados observados.</p> <p>El sistema de monitorización para gestión de alarmas desarrollado define dos clases, correspondientes a situaciones del proceso con o sin coque. Las características típicas de cada una de ellas se han determinado, en la primera fase del análisis, aplicando <i>clustering</i> borroso a los datos históricos almacenados sobre el funcionamiento de varios hornos. Una vez caracterizadas las clases, el sistema puede utilizarse para analizar el estado del proceso. Para cada situación examinada, el sistema proporciona al operador un único número por cada horno, que representa el grado de pertenencia de dicho horno a la clase que engloba los estados asociados a un exceso de coque (es decir, que requiere limpieza) permitiéndole planificar futuras actividades de mantenimiento.</p> <p>Cualquier proceso que requiera una monitorización puede supervisarse con un sistema de este tipo, que permite evaluar su estado clasificando la situación en un momento dado. Esta información facilita la organización y planificación del proceso, lo cual generalmente supone un ahorro significativo de costes.</p> <p><i>Este sistema ha sido desarrollado utilizando DataEngine. DataEngine es distribuido en España por DAEDALUS – Data, Decisions and Language, S.A. Nuestro departamento de Ingeniería está a su disposición para desarrollar proyectos como éste.</i></p>





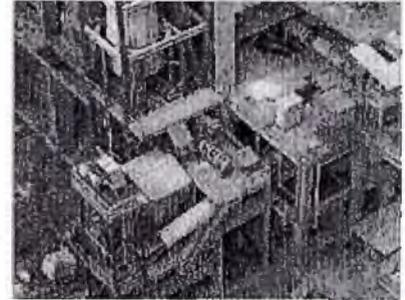
# Ingeniería de Datos en la Fabricación de Electrodomésticos

<b>NOMBRE DEL PROYECTO</b>	Aplicación de minería de datos en sistemas de control de calidad de electrodomésticos
<b>SECTOR</b>	Industria
<b>OBJETIVO</b>	Desarrollo de un sistema de control de calidad para detección de electrodomésticos defectuosos mediante análisis vibro-acústico.
<b>DESCRIPCIÓN</b>	<p>Según un estudio de mercado europeo, lavadoras y secadoras son, con diferencia, los electrodomésticos menos fiables, seguidos por los lavavajillas. La mayor parte de las pruebas a las que se someten estos productos a la salida de la cadena de producción comprueban los circuitos eléctricos, la seguridad o posibles fugas de agua. Sin embargo, las piezas mecánicas así como su correcto ensamblaje no son revisadas tan meticulosamente.</p> <p>Uno de los motivos es la dificultad para acceder a dichos componentes mecánicos. Para realizar este tipo de pruebas sería necesario desensamblar, al menos parcialmente, el producto, lo cual es inaceptable desde el punto de vista industrial. Los fabricantes necesitan un sistema de inspección automático que supervise de forma integrada en la línea de producción, y mediante métodos no invasivos, las características mecánicas del producto final. Los actuales métodos de revisión <i>on-line</i> de fallos mecánicos son escasamente eficaces, por su dependencia de un operador humano y por no resultar suficientemente fiables.</p> <p>Cualquier defecto mecánico de una máquina provoca la aparición de vibraciones anómalas. Sin embargo, dentro de los presupuestos industriales habituales es inviable el empleo de modelos matemáticos para representar la propagación de las vibraciones a través de la estructura de electrodomésticos, formados por más de un centenar de piezas interconectadas.</p> <p>El objetivo del sistema desarrollado en este proyecto es proporcionar una evaluación objetiva (independiente del operador) de la calidad vibro-acústica de la máquina. Para ello se ha diseñado un sistema de diagnóstico que analiza las señales medidas por una serie de sondas utilizando técnicas de minería de datos para identificar un conjunto de fallos mecánicos típicos.</p> <p>En este caso se ha construido un clasificador basado en redes neuronales que analiza los parámetros característicos de las señales medidas y determina, a partir de ellos, si la máquina en cuestión presenta algún fallo mecánico. Las redes neuronales permiten modelar sistemas complejos sin necesidad de que se conozca un modelo matemático de los mismos, obteniéndose con ellas resultados muy precisos, por lo que son una solución idónea para el tratamiento de problemas de este tipo.</p> <p>Este sistema ha sido instalado en la línea de producción de una importante empresa italiana y, dada su tasa de acierto en los diagnósticos, se espera que contribuya a reducir drásticamente las reclamaciones de los clientes.</p> <p><i>Este sistema ha sido desarrollado por MIT – Managment Intelligenter Technologien GmbH utilizando DataEngine V.i para LabView, como parte del proyecto MEDEA subvencionado por la comisión europea. DataEngine V.i es distribuido en España por DAEDALUS – Data, Decisions and Language, S.A. Nuestro departamento de Ingeniería está a su disposición para desarrollar proyectos como éste.</i></p>

<b>NOMBRE DEL PROYECTO</b>	Aplicación de minería de datos para predicción de calidad y optimización del proceso de producción de cemento
<b>SECTOR</b>	Industria
<b>OBJETIVO</b>	Desarrollo de un sistema de ayuda a la decisión para la optimización del proceso de producción en una fábrica de cemento.

## DESCRIPCIÓN

Debido a la complejidad y no-linealidad de la mayoría de procesos industriales, a menudo es difícil o incluso imposible establecer modelos matemáticos de los mismos. La minería de datos resulta especialmente útil para este tipo de problemas en los que no es posible aplicar métodos convencionales. Aplicando técnicas de minería de datos, en particular redes neuronales y algoritmos genéticos, se ha desarrollado un sistema que modela el proceso de molido del cemento, permitiendo predecir la calidad del producto a partir de las condiciones de operación, así como optimizar éstas para una determinada calidad.



En un molino de cemento, se tritura clinker, yeso y caliza. A la salida, un separador aparta el material que cumple los criterios de calidad respecto a finura, que se almacena en un silo, mientras que el resto del material se devuelve al molino. El sistema se refrigera mediante un circuito de ventilación y un sistema de inyección de agua, que disipan el calor generado por la fricción en el molino. La rotación del separador, la ventilación del separador y la inyección de agua en la cámara de pulverización son algunos de los factores que influyen en el proceso de molido.

Puesto que la trituración es la etapa final del proceso, tiene un impacto fundamental en la calidad final del producto, definida en este caso por el método de Blaine, que es una medida del área por unidad de masa de material granular. Por tanto, la cantidad y composición de los flujos de materiales así como la finura del producto a la salida deben controlarse y regularse cuidadosamente para cumplir con los criterios de calidad, que son continuamente supervisados. Utilizando redes neuronales se ha desarrollado un modelo que establece una correspondencia entre un conjunto de variables de estado y parámetros de control del proceso, y la finura de Blaine observada, lo que permite supervisar el estado de forma más precisa y, por tanto, mejorar el control del proceso. El modelo puede, entonces, utilizarse para determinar a priori los parámetros de calidad del cemento producido en el molino, a partir de las medidas del proceso.

El modelo permite además, aplicando algoritmos genéticos, calcular los parámetros del proceso que permitirán producir cemento de la finura requerida y con un coste mínimo. Este método puede servir como una herramienta de ayuda a la decisión para el operador del proceso, o para optimizar el proceso de forma automática.

La utilización de estas técnicas para modelado de procesos y determinación del punto de trabajo óptimo es una aproximación que no depende de características específicas del proceso, por lo que puede aplicarse de forma general, a cualquier problema análogo. Su interés se centra principalmente en dos campos de aplicación:

- Modelado de procesos (análisis condicional: "qué pasa si...").
- Optimización del punto de trabajo de acuerdo con un valor objetivo específico.

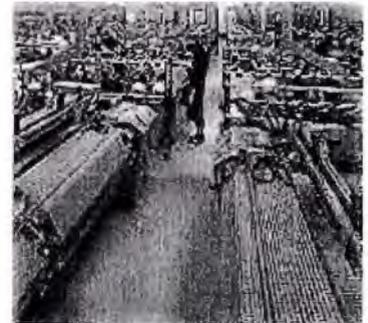
Mejorando el control de proceso con la ayuda de sistemas como el expuesto pueden conseguirse reducciones de coste significativas, basadas en menores tiempos de ajuste, óptima asignación de recursos y mejores estándares de calidad.

*Este sistema ha sido desarrollado por MIT – Managment Intelligenter Technologien GmbH y ENCI – Eerste Nederlandse Cement Industrie N.V. utilizando DataEngine. DataEngine es distribuido en España por DAEDALUS – Data, Decisions and Language, S.A. Nuestro departamento de Ingeniería está a su disposición para desarrollar proyectos como éste.*

<b>NOMBRE DEL PROYECTO</b>	Aplicación de minería de datos para la estimación de la resistencia mecánica de un material
<b>SECTOR</b>	Industria – textil
<b>OBJETIVO</b>	Desarrollo de un modelo que permita estimar la resistencia mecánica de un material

## DESCRIPCIÓN

La calidad de los materiales fabricados industrialmente se verifica mediante pruebas basadas en estándares normalizados. En la mayoría de los casos, estos tests se realizan en el laboratorio, fuera del proceso productivo. Para facilitar el control del proceso, los parámetros de calidad fundamentales deberían obtenerse durante la fabricación, sin embargo a menudo esto es imposible (por ejemplo, en el caso de pruebas destructivas). Como alternativa pueden medirse otras variables y calcular a partir de ellas, mediante un modelo adecuado, el parámetro de calidad buscado. No obstante, en muchas ocasiones no se dispone de dicho modelo porque se conocen sólo las principales variables de entrada, pero no las relaciones existentes entre ellas. Para este tipo de problemas, la minería de datos ayuda a encontrar un modelo que represente una buena aproximación de las relaciones.



En el caso de materiales producidos en rollos, como papel, textiles, láminas de metal o *no-tejidos (nonwovens)*, uno de los parámetros de calidad fundamentales es su resistencia mecánica. Es importante porque determina la fuerza que puede aplicarse al material en las siguientes etapas productivas y, por tanto, su rango de aplicaciones. En general, la resistencia se mide *off-line* utilizando equipos que determinan la fuerza a la que el material se rompe.

Determinar la resistencia según se está fabricando el material supondría una mejora sustancial del proceso, pero el procedimiento de medida estándar es destructivo, y no existe un método simple de medida durante el proceso productivo. Aplicando técnicas de minería de datos se ha desarrollado un modelo que permite estimar la resistencia mecánica a partir de otras medidas que pueden realizarse durante la fabricación.

La resistencia mecánica representa una propiedad dependiente de la elasticidad del material, característica que sí puede medirse durante el proceso productivo. Sin embargo, aunque ambos parámetros están estrechamente relacionados no existe correlación lineal entre ellos y la resistencia mecánica depende además de otras variables, destacando entre ellas el grosor del material y su densidad, que también pueden medirse *on-line*. No obstante, existen otros parámetros que también influyen en la resistencia del material y que no pueden medirse o para los que no se conoce un modelo.

Puesto que no se dispone de un modelo físico, se han aplicado técnicas de minería de datos para desarrollar un sistema que modele la relación de la resistencia con la elasticidad, la densidad y el grosor del material, en este caso *no-tejidos spunbound*. Analizando los datos registrados sobre los valores de estas variables y la resistencia medida en el laboratorio se ha configurado una red neuronal que proporciona una estimación de la resistencia. Esta estimación se obtiene a partir de los parámetros medidos durante el proceso de fabricación, sin necesidad de realizar pruebas específicas fuera de la cadena.

*Este sistema ha sido desarrollado por Freudenberg Vliesstoffe KG utilizando DataEngine. DataEngine es distribuido en España por DAEDALUS – Data, Decisions and Language, S.A. Nuestro departamento de Ingeniería está a su disposición para desarrollar proyectos como éste.*