

 UNIVERSIDAD
TORCUATO DI TELLA
MBA INTENSIVO 2010

MANTENIMIENTO INDUSTRIAL & DATA MINING

VALOR AGREGADO DE LA MINERÍA DE DATOS A LA TOMA DE DECISIONES

ALUMNO: FRANCISCO ASHFIELD

TUTOR: RAMIRO GALVEZ

LUGAR: BUENOS AIRES I ARGENTINA

00101100101110100011
00001000101011111011
10001010100110010111
00010111010101110011
01010110000011010110
00110100101011000110
11110 **Mantenimiento** 01100
1110011 **Industrial** 00010111
110101011 **y** 01010111100
111011 **Data Mining** 1011101
10110110100001110011
10111110011101101010
01000100011010100111
10101100010001010100
00010111001101000100
11100110111011011100

**Valor agregado de la minería de
datos a la toma de decisiones**

AGRADECIMIENTOS:

A Gretel, mi esposa, por su apoyo, colaboración y comprensión de siempre. Por la generosidad con la que me ha permitido dedicar a la tesis, un tiempo que le pertenece a ella y a nuestro pequeño Juan Manuel.

A la familia, amigos y compañeros, que de alguna manera han contribuido para que este trabajo pudiera llevarse a cabo.

Un agradecimiento muy especial a mi tutor Ramiro Galvez, por la excelente disposición con que siempre ha respondido a las reiteradas consultas, por las correcciones, aportes y orientaciones oportunamente realizadas, que han hecho posible alcanzar los objetivos buscados.

PALABRAS CLAVES

- Minería de Datos
- *Data Mining*
- Mantenimiento Industrial
- Procesos Industriales

ÍNDICE

INTRODUCCIÓN	p 6
SECCIÓN 1 - <i>Data Mining</i> (overview)	
1.1 ¿Qué es?, ¿Qué lo motiva?, ¿Para qué es importante?	p 8
1.2 Métodos de aprendizaje y técnicas de Data Mining.	p 13
1.2.1 Métodos directos: Supervisados.	p 13
1.2.2 Métodos indirectos: No supervisados.	p 15
1.3 El proceso cíclico de <i>DM</i> .	p 17
SECCIÓN 2 - Mantenimiento Industrial.	
2.1 Introducción a la problemática.	p 19
2.2 Un poco de historia. Estado actual del arte del mantenimiento.	p 21
2.3 Un problema de la vida real: “ <i>Data Sensor Chaos</i> ”.	p 27
SECCIÓN 3 - Mantenimiento Industrial y <i>DM</i>.	
3.1 Valor agregado del <i>DM</i> al Mantenimiento	p 31
3.2 Breve sondeo sobre el estado actual de aplicación de estas herramientas en mantenimiento.	p 43
SECCIÓN 4 - Dificultades. Requerimientos.	
¿Qué puede salir mal en un proceso de <i>DM</i> ?	p 45
SECCIÓN 5 - Conclusiones, consideraciones y recomendaciones futuras.	p 47
BIBLIOGRAFÍA.	p 48
ANEXO - Otras aplicaciones de <i>DM</i> en mantenimiento.	p 50

INTRODUCCIÓN

El Mantenimiento Industrial ha sufrido una gran cantidad de cambios en las últimas décadas, probablemente muchos más que otras disciplinas gerenciales. Estos cambios se deben no solo al gran aumento en variedad y cantidad de activos físicos que deben ser mantenidos, sino principalmente a una conciencia, compromiso e involucramiento del mantenimiento cada vez mayor con temas relacionados con seguridad, medioambiente y calidad de producto. Esto ha elevado al mantenimiento a una nueva dimensión. Nuevas aptitudes, calificaciones, maneras de pensar y de actuar son requeridas, tanto de ingenieros como de gerentes involucrados y responsables del mantenimiento.

Las nuevas industrias altamente automatizadas, con sistemas de control y supervisión cada vez más complejos, han hecho que como producto secundario de su operación, generen y dispongan de miles de datos constantemente. Sin embargo, esto ha puesto en evidencia una gran carencia: la gran dificultad que existe en el ámbito industrial para analizar grandes volúmenes de datos y poder obtener información útil para el negocio en tiempo y forma.

En particular existe un gran potencial en poder utilizar estos datos en mantenimiento, para poder predecir el estado de los equipos como ayuda o soporte a la toma de decisiones. En la mayoría de los casos, los datos o registros se suelen analizar luego de ocurrida la falla para ver cuando fue que se empezó a manifestar, lamentablemente tarde y sin haber evitado o disminuido sus consecuencias. Las empresas gastan millones en sistemas computacionales para gestión del mantenimiento (*CMMS*), por ejemplo *SAP-PM*, sin embargo son pocas las herramientas industriales disponibles para la integración y el análisis de los datos. Análisis que generalmente está constituido por herramientas estadísticas y planillas de cálculo, y fundamentalmente realizado a base del esfuerzo personal de ingenieros y planificadores.

El objetivo general de esta tesis de carácter descriptivo y exploratoria, será entonces demostrar la importancia del *Data Mining (DM)* para el mantenimiento industrial moderno, como una solución a esta dificultad anteriormente mencionada. Como objetivos específicos se busca acercar a la persona de mantenimiento industrial con la minería de datos, presentar sus ventajas, posibles aplicaciones prácticas, limitaciones, requerimientos y dificultades. Así como también presentar otra aplicación de *DM* quizás no tan divulgada.

No pretende ser una exposición o discusión formal sobre algoritmos de *DM*, ni sobre mantenimiento industrial, sobre estos temas existe mucha literatura y alguna se menciona en la bibliografía. Profundizar y discutir sobre cualquiera de estos tópicos escapa el nivel y el alcance del presente trabajo. El foco está en el valor agregado del *DM* para Mantenimiento, como herramienta y no en sus algoritmos o tecnología, que además de ser de una complejidad importante poco aportan al profesional de mantenimiento.

En la Sección 1 se realiza una breve descripción a modo de introducción de los objetivos que busca el *DM*, tipos de aprendizajes y las técnicas más usuales. Se busca explicar y clarificar, qué es y qué no es y mostrar los diferentes pasos de un proceso de minería de datos. En la Sección 2 se hace una introducción al problema del mantenimiento y una breve reseña del estado del arte del mismo con el objetivo de presentar dónde se está posicionado. Se presenta un caso real a modo de ejemplo para entender el problema planteado. En la Sección 3 se encuentra el foco de la presente tesis. En ella se considera el *DM* como una herramienta de gran valor para incorporar en el mantenimiento industrial. Para reforzar esta hipótesis se describen tres campos de aplicaciones prácticas en mantenimiento, identificando modos de fallas, prediciendo el mal funcionamiento de equipos, y aportando información útil sobre el estado de condición de los mismos. Se presenta además, una encuesta y un breve sondeo sobre el estado actual de aplicación de la minería de datos en mantenimiento. En la Sección 4 se presentan los requerimientos, las limitaciones y dificultades que los métodos de *DM* pueden encontrar en este campo de aplicación. Finalmente, en la Sección 5 se plantean las conclusiones, consideraciones y recomendaciones futuras.

SECCIÓN 1 - DATA MINING (OVERVIEW)

1.1 ¿Qué es?, ¿Qué lo motiva?, ¿Para qué es importante?

- ¿Qué es el DM?

“Data Mining is the task of discovering interesting and hidden patterns from large amounts of data where the data can be stored in databases, data warehouses, OLAP (online analytical process) or other repository information” [1].

“Data Mining involves an integration of techniques from multiple disciplines such as database technology, statistic, machine learning, neural networks, information retrieval, etc” [2]

“Data Mining is the exploration and analysis, by automatic or semiautomatic means, of a large quantities of data in order to discover meaningful patterns and rules”. [3]

“A business process for exploring large amounts of data to discover meaningful patterns and rules”. Aquí se agrega a la definición el concepto de proceso que interactúa con otros procesos, y que no tiene principio ni final, sino que por lo contrario se realimenta y genera más minería de datos. [4]

“Data Mining is a process of co-operation between human and computer. The best results can be obtained from the moderation between knowledge of experts in describing issues and the goals with computer searching capabilities” [5]

“Desde un punto empresarial, es la integración de un conjunto de áreas que tienen como propósito la identificación de un conocimiento obtenido a partir de las bases de datos que aporten un sesgo hacia la toma de decisión” [6]

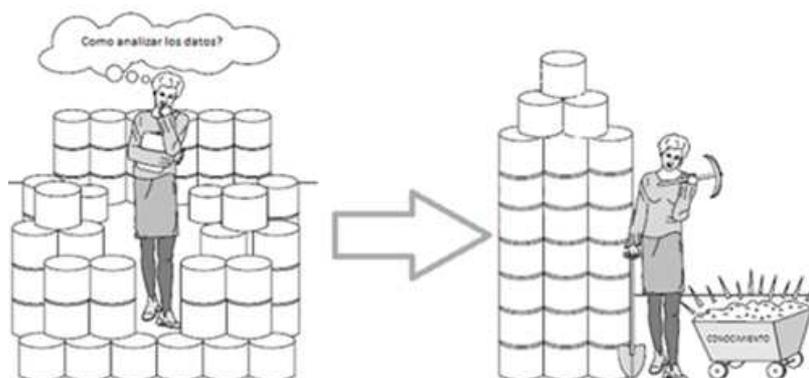


Fig. 1.1 La minería de datos permite obtener oro de una montaña de datos. [7]

Realmente existen muchas definiciones de *DM*, siendo comúnmente la más aceptada, como un conjunto de pasos analíticos diseñado para la extracción de patrones de información (implícitos, no triviales, desconocidos y potencialmente útiles) a partir de grandes cantidades de datos. A lo que se agrega el concepto de Proceso Cíclico de Negocio [4], que se detallará más adelante.

- ¿Qué no es la minería de datos?

- No es una varita mágica, no sustituye la inteligencia humana ya que es necesaria para el proceso de *DM*.
- No es solo análisis estadístico.
- No es un sistema experto o *DSS (Decision Support System)*
- Encuentra patrones o relaciones pero no establece su valor agregado para el negocio (Ejemplo, no discrimina trivialidades, ni incoherencias).
- No infiere causalidad. Por ejemplo, podría encontrarse que ciertas personas con un cierto nivel de ingresos y que hacen ejercicio regularmente, tienen mayor probabilidad de comprar un determinado producto, sin embargo, esto no significa que esos factores causan la compra del producto, sólo que la relación existe.

- ¿Qué motiva el *DM*? “La necesidad es la madre de la innovación” – Plato (400 A.C)

Se puede resumir que el auge actual del *DM* se ha potenciado básicamente por la convergencia de 5 factores:

a) Competencia comercial: Fuerte competencia comercial y la necesidad de generar ventajas competitivas entre empresas ricas en información: Seguro, *retail*, servicios financieros. La información como producto en sí misma.

b) Producción de datos: *DM* sólo tiene sentido cuando se dispone de grandes volúmenes de datos: *scanners* de venta, tarjetas de crédito, *pay per view*, telecompra, sistemas de información industriales, etc.

c) Almacenamiento de datos: Los datos se están extrayendo de los sistemas operacionales para ser guardados en almacenes de datos (*Data Warehouse*) que contienen la memoria de las corporaciones. El almacenamiento de datos es el proceso de reunir datos dispares de toda la organización para fines de apoyo a la decisión.

d) Capacidades computacionales: Los algoritmos de minería de datos suelen requerir varias pasadas de grandes cantidades de datos, por lo que son computacionalmente intensivos. La continua disminución de precios de discos, memorias, potencia de procesamiento y ancho de banda de las redes, ha traído al alcance de los negocios comunes, costosas técnicas que tradicionalmente se reservaron sólo a unos pocos laboratorios financiados por los gobiernos.

e) Disponibilidad y accesibilidad de programas comerciales: Siempre existe un desfase entre que una innovación es presentada en un ambiente académico, y el momento que ésta es aplicada comercialmente en algún software. También existe un desfase entre este primer

producto y el tiempo en el cual se logra una amplia aceptación de los resultados del mismo en el mercado. Para el *DM* este periodo ya ha expirado, hoy en día se encuentran disponibles numerosos programas, algunos comerciales y otros de uso libre. ¹

- ¿Para qué es importante el *DM*?

Para ver su importancia lo mejor es presentar algunos casos reales exitosos de empresas muy conocidas, que han mejorado sus procesos de negocios y obtenido grandes beneficios económicos gracias a la minería de datos. Los casos seleccionados que se presentan a continuación, ilustran una variedad de aplicaciones que van desde la evaluación de la solvencia de los prestatarios, mejorar la eficiencia del marketing directo, los ajustes de precios en tiempo real, detección de fraudes y la gestión de la cadena de suministro.

Capital One-Financial Group

Las compañías de servicios financieros son las que mayormente han adoptado tecnologías de *DM*. *Capital One Corporation* fue pionera en la toma de decisiones en base al análisis de grandes conjuntos de datos en sus operaciones crediticias. En el pasado, las tarjetas de crédito eran emitidas sólo a aquellos clientes que tenían un historial sólido de reembolso, y los préstamos eran otorgados siempre que los prestatarios pudieran presentar garantías firmes. *Capital One* cambió las reglas del juego mediante el uso de los datos para determinar la capacidad crediticia de sus clientes. Su estrategia se basaba en la premisa de que los clientes, sin un historial de crédito, podrían tener el potencial para ser buenos prestatarios. Un ejemplo típico es un graduado universitario, con recursos limitados y una deuda estudiantil, pero con potencial para hacer una exitosa carrera empresarial. Es así que *Capital One* creó un almacén de datos enorme, un equivalente de 20 páginas aproximadamente para cada uno de sus 30 millones de clientes, con información relevante para los préstamos sub-prime. Utilizó esta información para llevar a cabo experimentos, hasta 45.000 pruebas de producto y marketing fueron hechas en el 2001. Se encontraron ofertas financieramente sólidas para sus diversos segmentos. En resumen con ayuda del *DM*, *Capital One* pasó de 1 a 12,8 billones de préstamos gestionados en 8 años. [5]

¹ (Ejemplo: <http://www.cs.waikato.ac.nz/ml/weka/>)

Victoria 's Secret – Ropa y elementos de belleza femenina

La empresa tuvo un rápido crecimiento en sus ventas cuando lanzó su sitio web en 1998. Al mismo tiempo, las ventas por catálogo se desplomaron. La empresa tuvo que replantear su estrategia de canal, especialmente porque los costos de atender a los clientes por los catálogos impresos eran más altos y más clientes fueron adquiridos por la web. No obstante la compañía todavía creía que era necesario distribuirlos como complemento a su canal web, y debía buscar la manera de dirigir sus catálogos de la forma más eficiente. La distribución de éstos se dirigió entonces a los clientes basados en *RFM* (reincidencia, frecuencia y valor monetario). Los clientes con puntuaciones más altas en *RFM* son más propensos a recibir catálogos. Esto ayudó a la compañía a reducir su distribución de catálogos de un máximo de 400 millones anuales entre 1978 y 1998, a una cantidad de 350 millones entre 1998 y 2002.²

Harrah 's Entertainment

Los operadores de casinos suelen ofrecer tarifas bajas de hotel para motivar a sus clientes a gastar más en los juegos. Sin embargo, no todos los clientes gastan lo suficiente para justificar los descuentos en las tarifas del hotel.

Harrah 's Entertainment, uno de los mayores operadores de casinos del mundo, utiliza la minería de datos para analizar el gasto en tiempo real de los juegos, y les ofrece incentivos mientras todavía se encuentran hospedados en el hotel. Procesa 16 terabytes de datos cada día que proceden fundamentalmente de sus máquinas de slots. Con estos datos logra analizar el tiempo dedicado a cada uno de los juegos, las preferencias de género y una gran cantidad de otras variables sobre las preferencias de sus clientes. En consecuencia la proporción del gasto en los juegos aumentó de 36% a casi el 50% después de que comenzaron a utilizar las tecnologías de minería de datos.³

VISA

El fraude con tarjetas de crédito tiene una tasa endémica del 0,93% para operaciones offline y una más alta aún de 1,97% para operaciones online. El desafío de detectar fraudes es evitar las falsas alarmas. En el pasado, 97 falsas alarmas eran generadas por cada alarma genuina. Las compañías de tarjetas de crédito tenían que encontrar el punto de equilibrio entre mantener el índice de fraudes suficientemente bajo y no irritar a los clientes con falsas alarmas. Esto es difícil de conseguir con modelos estadísticos que comparan el comportamiento normal, con un

² Fuente: <http://www.tuck.dartmouth.edu/cds-uploads/case-studies/pdf/6-0014.pdf>

³ Fuente: <http://www.forbes.com/forbes/2004/1213/110.html>

comportamiento anormal porque los criminales son lo suficientemente astutos para sortearlos.

VISA presenta entonces un nuevo software (*Visor*), puntuación VISA inteligente del riesgo, a través de todos los bancos en Europa para reducir la tasa de fraude. Su nuevo sistema analiza el comportamiento anormal de no sólo el titular de la tarjeta, sino también de cada comerciante. Además, ahora se utiliza software de inteligencia artificial que cambia las reglas de identificación de fraude basado en los datos más recientes. La tasa de fraude se redujo de 1.576 a 458 casos, y las falsas alarmas se han reducido por diez. [8]

Wal-Mart

Las tiendas minoristas luchan por equilibrar su inventario con la demanda de los consumidores, y sus proveedores muchas veces son incapaces de enviar las cantidades correctas de los productos a los almacenes en tiempo y forma. A raíz de esto *Wal-Mart* creó un almacén de datos masivo en la década de 1990 para almacenar información sobre las ventas y los inventarios. Su programa *Retail Link* comparte información con sus proveedores como *Procter and Gamble* por ejemplo, quién tiene además mucha experiencia en la predicción de demanda del consumidor. La disponibilidad de información en tiempo real ha permitido a los proveedores almacenar mercadería sin esperar a que lleguen las órdenes de compra para ser colocadas. Wal-Mart además, es capaz hoy en día de responder rápidamente a circunstancias totalmente excepcionales de demanda, como pueden ser por ejemplo las generadas por las tormentas en la Florida. Con base en experiencias pasadas, puede predecir cuál demanda de un determinado producto va a aumentar y notificar a sus proveedores.⁴

⁴ Fuente: *What Wal-Mart Knows*

AboutCustomers'Habits<http://www.nytimes.com/2004/11/14/business/yourmoney/14wal.html>

1.2 Métodos de aprendizaje y técnicas de DM.

DM & Estadística

Si bien la minería de datos incluye herramientas de estadística, su enfoque es muy diferente. La estadística y los métodos de consultas de base de datos deductivas se basan en hipótesis que a priori deben ser conocidas o sospechadas y que se buscan demostrar o comprobar. Son dirigidos por hipótesis (*Hypothesis Driven*). La minería en cambio no conoce a priori estas relaciones o hipótesis y las conclusiones son obtenidas a partir de los datos directamente. Son métodos conducidos por los datos (*Data Driven*) establecen relaciones, agrupaciones y/o asociaciones entre los datos, no establecen relaciones de causa - efecto.

Es justamente este carácter *Data Driven* de la minería, lo que la hace tan poderosa para analizar y obtener conclusiones en aquellos problemas donde establecer relaciones entre los atributos y/o variables no sea una tarea fácil o sea prácticamente imposible, por la cantidad o complejidad de los mismos.

Como ya fue mencionado anteriormente, no es el objetivo de este trabajo entrar en el detalle de los diferentes métodos de aprendizaje y sus técnicas. Sin embargo, se puede decir que la vasta mayoría de métodos de aprendizaje automático se engloban dentro de dos grandes clases: Directos o supervisados, y los indirectos o no supervisados. Algunas de las técnicas más comunes de ambos métodos se describen también a continuación.

1.2.1 Métodos directos: Supervisados

Estos métodos estiman el valor de un atributo denominado objetivo o etiqueta, de un conjunto de datos, conocidos otros atributos llamados descriptores. A partir de un conjunto de datos cuyo atributo etiqueta se conoce, se induce una relación entre dicha etiqueta y la otra serie de atributos. En los métodos supervisados diferentes variables juegan diferentes roles. Las variables objetivo son los objetos de estudio, las otras son usadas para explicar o predecir el valor de las variables objetivo.

En otras palabras el *DM* directo no busca tan solo patrones en los datos, sino que busca patrones que expliquen los valores de los atributos objetivos. Esas relaciones sirven para realizar la estimación de datos cuya etiqueta es desconocida. Esta forma de trabajar se conoce como aprendizaje supervisado y se desarrolla en dos etapas:

- Entrenamiento: construcción de un modelo usando un subconjunto de datos con etiqueta conocida (*Training Set*)
- Prueba: Verificación del modelo sobre el resto de los datos (*Test Set*)

Las técnicas supervisadas se pueden dividir en⁵:

- Classification: El resultado a predecir es de clase discreta, (árboles de decisión, vecinos más cercanos, regresión logística, SVM, Naïve Bayes, redes neuronales, etc). Una de las técnicas usuales es el C5.0.
- Regression: El resultado a predecir es de carácter continuo numérico, (regresión lineal simple, SVR, árboles de regresión, redes neuronales, análisis de series de tiempo, etc).

La Fig. 1.2 muestra un ejemplo del resultado obtenido por un método de clasificación y representado como regla de clasificación, en este caso aplicado a la clasificación de clientes según el riesgo crediticio que estos representan [9]

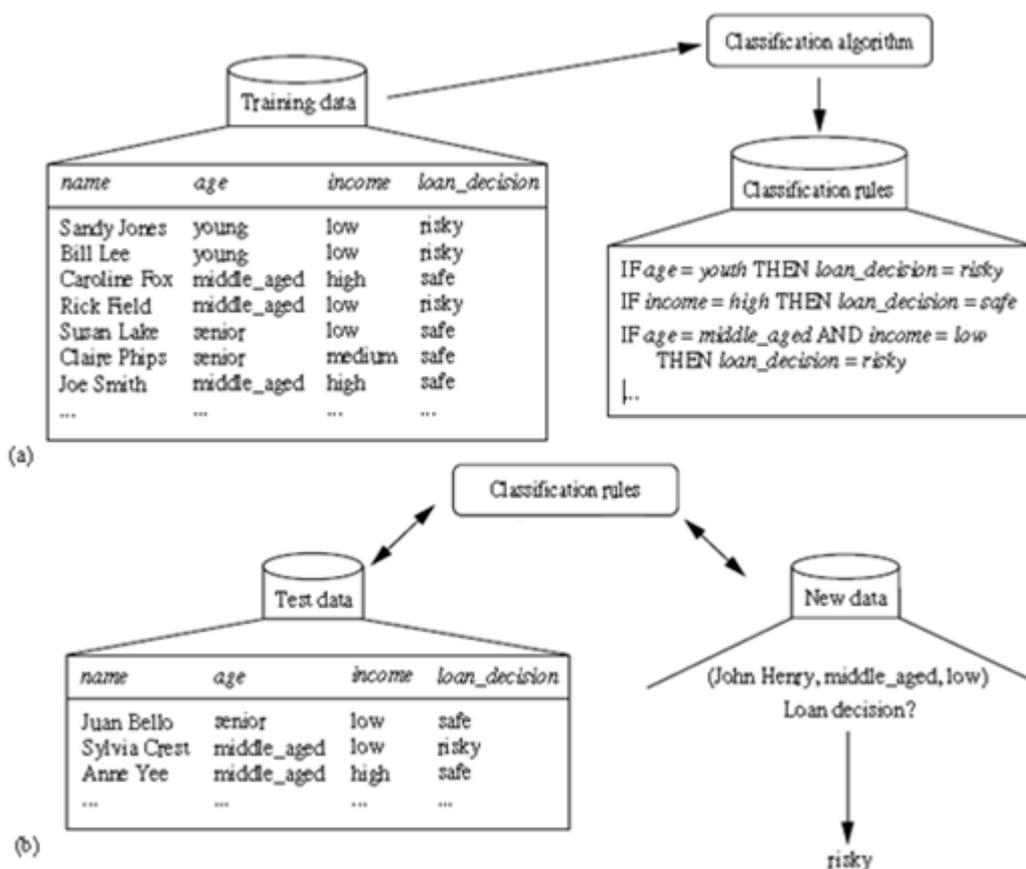


Fig 1.2 El proceso de clasificación de datos:(a) Entrenamiento: los datos de entrenamientos son realizados por un algoritmo de clasificación de *DM*. Aquí el atributo de clase o etiqueta es *loan_decision*, y el modelo aprendido o clasificado es representado en forma de regla de clasificación (classification rule); (b) Clasificación: datos de prueba son utilizados para estimar la precisión de la regla de clasificación. Si la precisión es considerablemente aceptable, la regla puede ser aplicada para la clasificación de nuevos datos.

⁵ Una técnica puede adaptarse para trabajar tanto como un clasificador así como una regresión.

Otro ejemplo, si se dispone de una base de datos con clientes con los siguientes atributos: *age* (*Youth, middle-aged, senior*); *student* (*yes, no*); *credit-rating* (*fair, excellent*). Mediante un algoritmo de clasificación y representado el resultado mediante un árbol de decisión podemos obtener un resultado similar al de la Fig. 1.3. Las hojas finales del árbol representan los atributos de clase (compra-computador = si ó compra-computador = no).

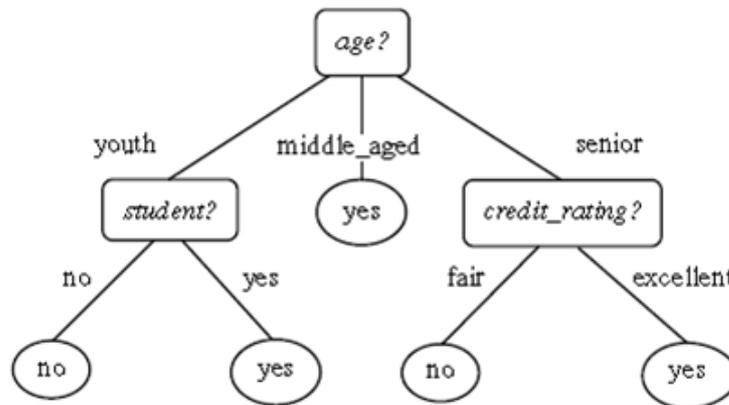


Fig. 1.3 Árbol de decisión para el concepto compra-computador

1.2.2 Métodos indirectos: No supervisados

Los métodos no supervisados a diferencia de los supervisados no utilizan una variable como objetivo o etiqueta. El método en si consiste en encontrar patrones totales, las variables no tienen un rol especial. Son especialmente útiles cuando queremos buscar clases, agrupaciones, semejanzas entre datos dentro un conjunto que inicialmente son desconocidas e incluso totalmente insospechadas. Después que un patrón es descubierto es total responsabilidad de la persona interpretarlo y decidir qué tan útil es.

Algunas de las técnicas de los métodos supervisados más comunes son:

- *Clustering*: Agrupa los datos en conglomerados o *clusters*. Es una técnica de obtención de patrones descriptivos, que busca encontrar grupos de observaciones homogéneas en los datos, denominados también modelos de segmentación. La meta de estas consultas, es tomar un conjunto de entidades representadas por registros de una base de datos, y dividir las en grupos o *clusters* de manera tal que los elementos dentro de cada grupo sean similares y distintos a la de los otros grupos. Una de sus técnicas usuales es *Kmeans*.

Una aplicación típica puede ser una investigación de marketing, donde se mide una serie de variables relacionadas con el comportamiento del consumidor para una amplia muestra de encuestados. El propósito del estudio es detectar "los segmentos del mercado", es decir, grupos de encuestados que de alguna manera son más similares entre sí (y los demás

miembros del mismo grupo) en comparación con los encuestados que “pertenecen a otros grupos”. Además de identificar estas agrupaciones, también es útil para determinar qué tan diferentes son estos grupos, cuales son las variables específicas o dimensiones en la que los miembros de diferentes agrupaciones pueden variar, y cómo.

➤ Association Rules: las reglas de asociación se utilizan para descubrir hechos que ocurren en común dentro de un determinado conjunto de datos. Asume que todos los datos son categóricos. No es una buena técnica para datos numéricos. El uso tradicional ha sido en Market basket analysis (Fig. 1.4). El tipo de resultado o relación que arrojan es del tipo: (leche, pan) → manteca. Vale la pena volver a resaltar que la flecha no implica causalidad, sino simplemente que la relación existe.



Fig. 1.4 Análisis de la canasta de mercado

➤ Técnicas de visualización: Evidentemente siempre la representación gráfica ayuda a una mejor interpretación de resultados. Inclusive en los problemas multidimensionales es posible representar los datos de dimensión $n > 3$ en gráficos de dimensiones menores, en particular en 2 y 3 dimensiones (Ej. *Sammon Mapping*). Esto ayuda mucho por ejemplo para ver la existencia de *clusters* y reglas de asociación (Fig. 1.5)

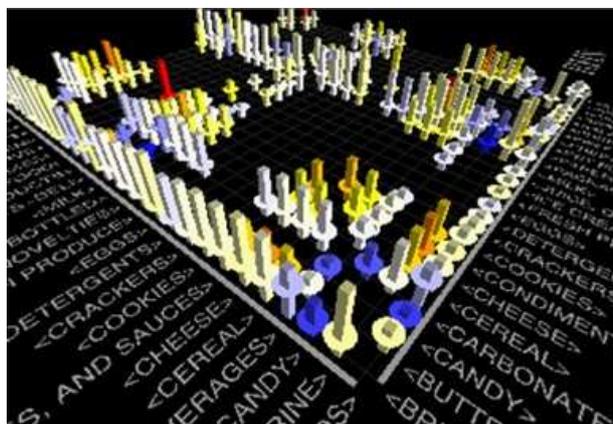


Fig. 1.5 Visualización de reglas de asociación en un supermercado [9]

1.3 El proceso cíclico de DM

En la literatura de la minería de datos, diferentes "marcos generales" se han propuesto para servir como modelos de organización para el proceso de recopilación de datos, análisis de los mismos, la aplicación de los resultados, y el monitoreo del proceso de mejora.

Algunos de estos marcos con sus principales pasos son:

- *CRISP-DM*, (*Cross Industry Standard Process for Data Mining*). Es una de las metodologías más populares que utilizan los mineros expertos para abordar los problemas. Principales pasos: Entendimiento del negocio, selección de los datos, preparación de los datos, modelado, evaluación e implementación.
- *DMAIC*, metodología *Six Sigma* introducida por Motorola en los 80s con el objetivo de mejorar la variabilidad de los procesos. Principales pasos: Definir (*Define*) el problema, medir (*Measure*), analizar (*Analyze*), mejorar (*Improve*) y controlar (*Control*).
- *SEMMA* (*SAS Institute*), centrado más en aspectos técnicos de la minería de datos. Etapas: muestreo (*Sample*), explorar (*Explore*), modificar (*Modify*), modelar (*Model*) y evaluar (*Assess*).

DM & Business

Con centro en la aplicación del *DM* como valor agregado a los negocios, resulta útil la metodología denominada como Ciclo Virtuoso del *DM* [4], que reconoce los siguientes pasos (Fig. 1.6):

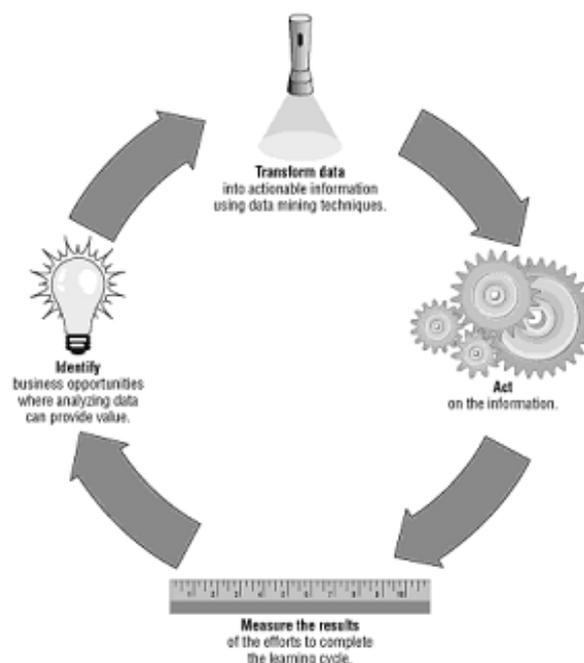


Fig. 1.6 El ciclo virtuoso del *DM* [4]

- **Identificar** una oportunidad de negocio donde aplicar *DM* puede agregar valor al negocio.
- **Transformar** los datos en información utilizable. Este paso es crítico y es generalmente el más largo. Consiste en la selección de los datos, limpieza o depuración, eliminar ruido pre-procesamiento en general y transformación de los datos.
- **Actuar** sobre la información para obtener una ventaja competitiva y mejorar nuestro negocio.
- **Medir** los resultados adecuadamente sobre los esfuerzos realizados.

El proceso de *DM* es un proceso cíclico, se obtienen resultados y en función de ellos se comienza una nueva interacción, en donde las salidas anteriores y el aprendizaje obtenido sirven para mejorar o corregir el proceso de *DM*. De esta forma, en cada ciclo se obtiene más cantidad de información y de más calidad para el negocio.

SECCIÓN 2 - MANTENIMIENTO INDUSTRIAL

2.1 Introducción a la problemática

“La situación en que estamos es como tener una represa que se está llenando de agua, cada vez más llena, y estamos tratando de sacar agua de ella con una pajita”.⁶

Producto de su operación las modernas plantas Industriales actuales procesan, generan y registran constantemente un gran volumen de datos. Ejemplo de estos son lotes de producción, niveles de tanques, velocidades de máquina, toneladas producidas, caudales, presiones, temperaturas, intervenciones de mantenimiento, niveles de vibraciones mecánicas, medidas de PH, conductividades, turbidez, corrientes eléctricas, potencias consumidas y/o generadas, emisiones ambientales, etc. Paradójicamente en esta sociedad de la información, con un crecimiento estimado de los datos almacenados que se duplica cada 20 meses (Fig. 2.1), y capacidades de ordenadores, medios de almacenamiento, y velocidades de comunicaciones que avanzan explosivamente; la capacidad de los profesionales de mantenimiento de analizar datos y obtener información no lo ha hecho de la misma forma. El manejo de bases de datos y técnicas estadísticas tradicionalmente bien establecidas en ingeniería, no están resultando suficientes ante la complejidad de las nuevas máquinas, de los sistemas de adquisición, los avances en tecnología de la información y sobretodo la gran producción de datos.

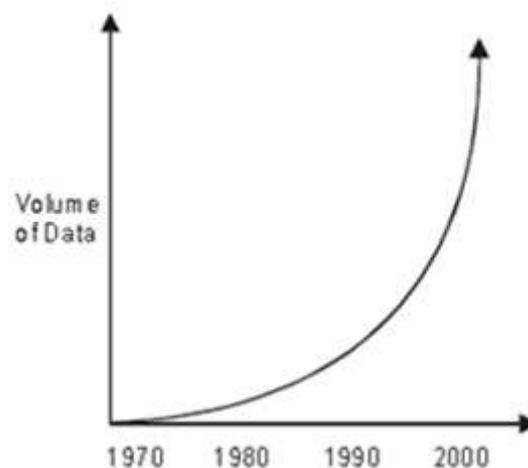


Fig. 2.1 Creciente Base de datos.

Si a esto se le agrega que para tener precisión y correcta interpretación se deben analizar simultáneamente varios y a veces cientos de objetos, cada uno con decenas de atributos además, entonces llegamos a la conclusión de que se necesitan nuevas herramientas, más automáticas y menos dependientes de las capacidades individuales de expertos (cada vez más

⁶ James Gray, Manager of Microsoft's Bay Area Research Center

escasos y costosos por cierto). Se podría resumir que existe un desfase entre la cantidad de datos disponibles y las capacidades de análisis o herramientas utilizadas en el mantenimiento industrial. Algo más preocupante que carecer de datos es tenerlos y no saber cómo aprovecharlos.

Los costos de mantenimiento pueden ascender hasta un 30% de los costos de producción dependiendo de la industria. Existe entonces una gran oportunidad económica en poder administrar y convertir estos datos en información útil, para predecir el estado de los equipos evitando fallas y costosas paradas inesperadas. Con esta información, el técnico de mantenimiento genera conocimiento que le permite optimizar sus recursos y realimentar sus planes y/o rutas de inspección, mejorando su performance. Se puede resumir que el "arte" actual del Mantenimiento es tomar acciones preventivamente *"just in time"*. Si éstas se realizan con demasiada frecuencia o ineficazmente, importantes y costosos recursos son desperdiciados (sobremantenimiento). Por otro lado, si el equipo se degrada por debajo del punto de fallo costosas paradas no planificadas pueden ocurrir. En resumen el objetivo será buscar el punto de equilibrio entre los costos de mantenimiento y los producidos por pérdidas de producción debidas a una falla (Fig. 2.2)

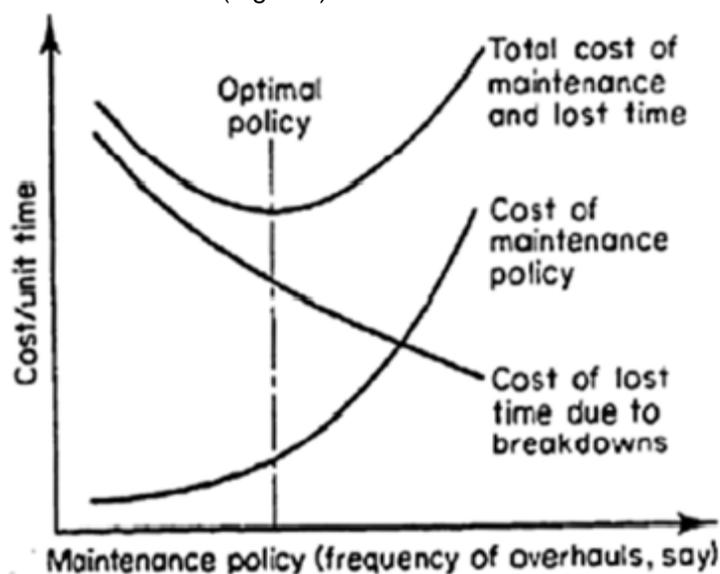


Fig. 2.2 Relaciones de los costos de mantenimiento

En conclusión, un objetivo constante de mantenimiento es la búsqueda de métodos para poder bajar los costos sin sacrificar disponibilidad de los equipos. Se entiende por disponibilidad de un equipo, al tiempo en que este se encuentra disponible para producción igual al tiempo total menos el tiempo de mantenimiento, dividido el tiempo total. Las modernas plantas dentro de sus *KPI (key process indicators)* esperan disponibilidades del 98% o más. En función de los porcentajes del costo de producción que represente el mantenimiento en la organización, cualquier herramienta o innovación que contribuya a disminuirlos, aunque sea en un pequeño porcentaje, por supuesto sin afectar su performance, puede significar millones de dólares en ahorro de costos de producción.

2.2 Un poco de Historia. Estado actual del arte del Mantenimiento [10]

El mantenimiento industrial ha sufrido varios cambios de filosofía desde la década del 40 hasta la fecha (Fig. 2.3). La primera generación conocida como producir hasta que rompa (*RTF, Run to Failure*), las gerencias no justificaban demasiado invertir en mantenimiento, las máquinas eran poco mecanizadas, muy simples, con fallas fáciles de reparar, y generalmente estaban sobredimensionadas, por lo tanto menos proclives a fallar, eran más confiables. Pocas habilidades eran requeridas de mantenimiento, básicamente lubricación y limpieza. Si se rompía simplemente se arreglaba.

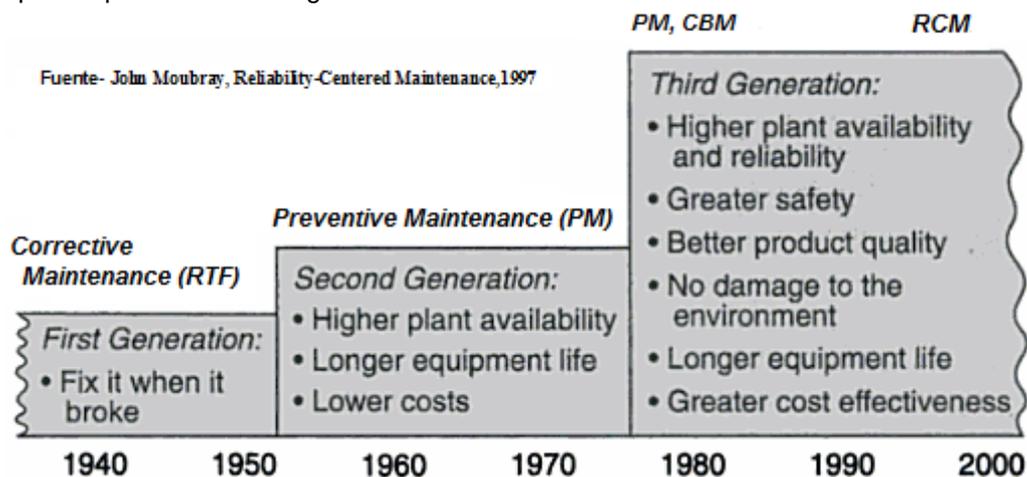


Fig. 2.3 Expectativa creciente del Mantenimiento

Posteriormente con los tiempos de la Segunda Guerra Mundial se produce un aumento de la demanda de todo tipo de bienes y conjuntamente una disminución significativa de la mano de obra de trabajadores industriales. Esto trajo como consecuencia la necesidad de desarrollo de maquinaria más compleja, y una mayor dependencia industrial de las mismas. Aumenta el grado de especialización de los técnicos de mantenimiento, el costo de la mano de obra, la importancia de la prevención de fallas, y surge el concepto de pérdida teórica de producción asociada con el costo de oportunidad de no producir. El tiempo de parada de máquina afecta la capacidad de producción de los activos físicos, aumenta los costos operacionales y sobretodo perjudica el servicio al cliente.

Observación: el costo de una reparación no planificada puede llegar a ser más de tres veces la de una parada bien planificada [11], básicamente por un tiempo de parada (*downtime*) más largo por falta de disponibilidad de recursos.

Evitar fallas resulta entonces crucial, y surge la idea del Mantenimiento Preventivo (*PM*), la Segunda Generación del mantenimiento muy popular hasta comienzo de los 90. El *PM* básicamente creía que la probabilidad de falla estaba correlacionada con el envejecimiento, y en consecuencia se limitaba a cambiar componentes y/o intervenciones periódicas prefijadas. En esta época se suponía que la mayoría de los equipos y/o máquinas tenían una curva de probabilidad de falla conocida popularmente como curva de la bañera, (alta probabilidad de

falla al comienzo y al final de la vida útil, en el medio un período de probabilidad de falla relativamente baja). La idea era entonces actuar antes que la probabilidad comenzara a aumentar. El periodo con alta probabilidad al comienzo se atribuía a la "mortalidad infantil" de los equipos (expresión usada en mantenimiento para aquellos componentes que fallan prematuramente luego de su puesta en marcha por primera vez).

Como consecuencia de esta política de recambio periódico, el costo de mantenimiento comienza a crecer rápidamente con relación a los operacionales, y esto obliga a una mejor planificación y control del mismo. Es así que a mediados de los años 90 con el objetivo de maximizar beneficios, sin perder foco en la confiabilidad y reducir costos de mantenimiento, adquiere importancia como concepto, el Mantenimiento Basado en la Condición (*CBM*), también conocido como Mantenimiento Predictivo (*PdM*). El cambio de componentes en forma periódica y fija ya no es una regla, sino que para cada caso se realiza un estudio del modo de falla y se ejecuta el mantenimiento más conveniente (adquiere fuerza el *RCM*, *Reliability Center Maintenance*)

Mantenimiento centrado en la confiabilidad. (*RCM*)

El *RCM* conocido como mantenimiento centrado en la confiabilidad, es la última filosofía en mantenimiento, tiene como objetivo ser costo eficiente, aplicando la estrategia de mantenimiento más adecuada a partir de un estudio sistemático de los modos o formas en que puede fallar un equipo, y con el objetivo de que éste continúe trabajando con las prestaciones requeridas o esperadas de su función, muchas veces inclusive menores a la nominal del equipo. Mantenimiento basado en la función y no en el equipo. A modo de ejemplo, supongamos que tenemos un auto que es capaz de alcanzar 200 km/h, sin embargo nosotros no precisamos que funcione a más de 140 Km/h, entonces haremos las tareas de mantenimiento para que pueda andar a 140 Km/h y no gastaremos un solo peso más.

En el *RCM* convergen las tres generaciones del mantenimiento, *RTF*, *PM*, y el *PdM* o *CMB*. Los equipos presentan diferentes curvas de probabilidad de falla en el tiempo y la actividad de mantenimiento debe ajustarse a estas características. Por ejemplo, ciertos componentes (generalmente electrónicos) tienen alta probabilidad de falla al comienzo, en el periodo denominado de mortalidad infantil, entonces realizar el cambio prematuro no solo no mejora la confiabilidad, sino que aumenta la probabilidad de falla y los costos de mantenimiento. La Fig. 2.4 muestra los distintos puntos de vista de los patrones de falla de los equipos según las distintas épocas y las técnicas de mantenimiento respectivas.

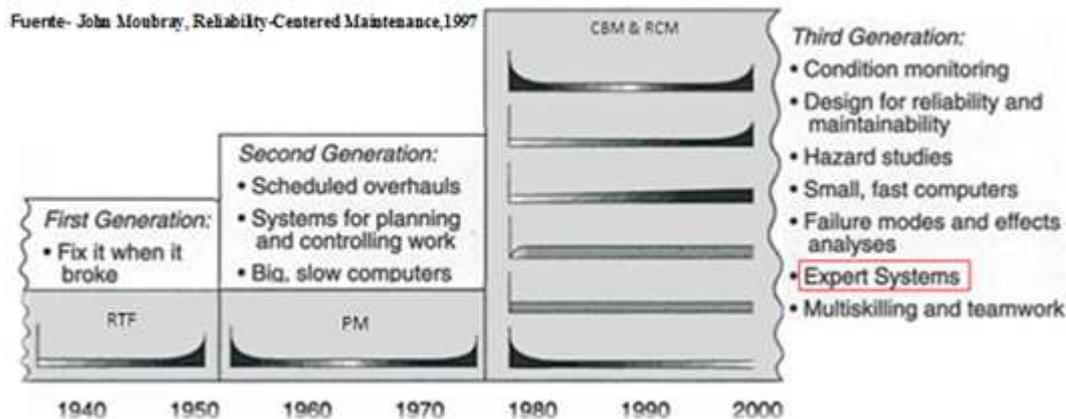


Fig. 2.4 Cambios en los puntos de vista de las fallas de los equipos y las técnicas de mantenimiento.

Resumiendo, la Fig. 2.5 representa la política de mantenimiento modernamente más aceptada. Conviven las tres filosofías pero de una forma mucho más eficiente, dependiendo la política empleada de las características y modos de fallo del activo físico. Si bien el gran protagonista es el Mantenimiento Basado en la Condición, se puede ver que aún hoy algunos equipos continúan manteniéndose en forma correctiva o emergencial (*RTF*). Ya sea porque no hay manera de predecir su fallo o porque es más económico así hacerlo. Política generalmente aplicada para componentes no críticos, cuyo fallo no representen riesgos para el proceso, la seguridad o el medioambiente, y donde realizar inspecciones o monitorear su condición tiene poco valor agregado a producción o alto costo (Ejemplo: un artefacto de iluminación). Lo mismo ocurre para aquellos componentes que se puede predecir su vida útil, se cambian según periodicidad establecida y no se invierte en monitoreo o inspección. (Ejemplo: correas de una polea).

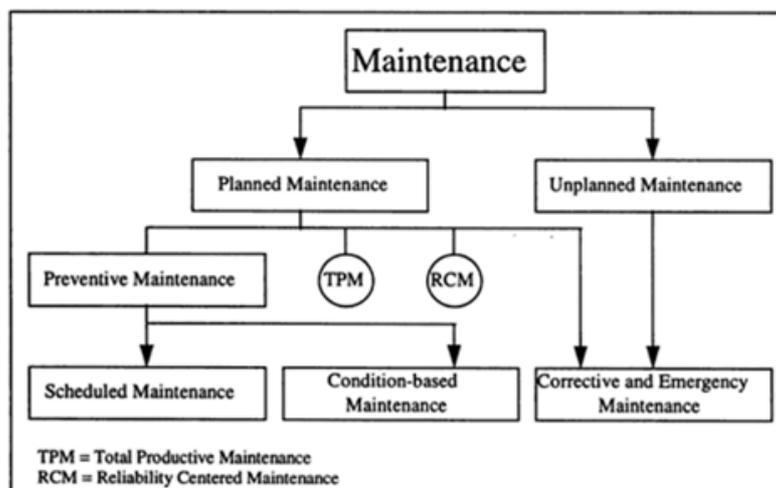


Fig. 2.5 Políticas de mantenimiento

El gran objetivo de un programa de mantenimiento basado en la condición (*CBM*), es conocer el estado de la maquinaria, de tal manera que se pueda determinar su disponibilidad de manera segura, eficiente y económica. Estas técnicas están dirigidas a la medición de variables físicas, indicadores de la condición de la máquina, y mediante un análisis adecuado, efectuar la comparación con valores normales, para así determinar si está en buen estado o en condiciones de deterioro. Esta estrategia asume que hay características medibles y observables que son indicadores de la condición de la maquinaria.

Objetivos del *CBM*:

- Detectar condiciones que pueden ser causa de falla – (proactiva)
- Detectar problemas en la maquinaria – (predictiva)
- Evitar fallos catastróficos – (predictiva)
- Diagnóstico de la causa de la falla – (proactiva)
- Pronóstico de utilidad – (predictiva)

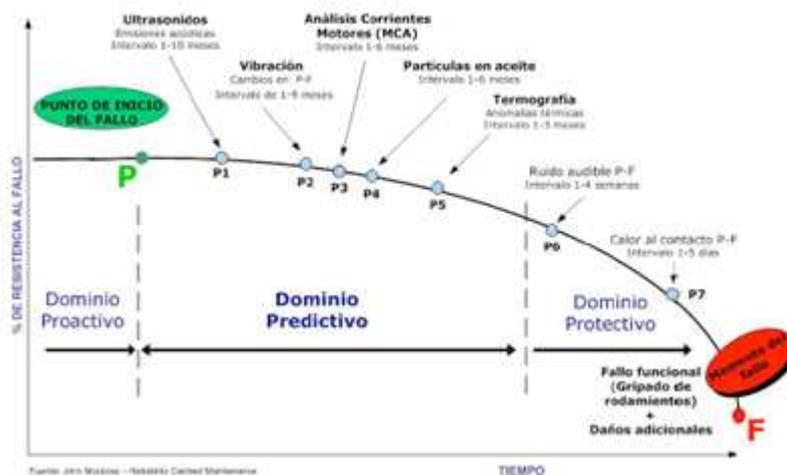


Fig. 2.6 Técnicas a utilizar en el dominio predictivo según modo de fallo

El monitoreo de condición estudia la evolución de los parámetros seleccionados en función del tiempo, y establece una tendencia que indica la existencia de un fallo, su gravedad y el tiempo en que el equipo puede fallar. La toma de decisiones a tiempo permite evitar que el fallo se produzca, eliminando la posibilidad de un fallo catastrófico. La ventaja de esta estrategia es que el estado de condición puede ser efectuado mientras el equipo está funcionando, y de esta manera se planifican las acciones correctivas oportunamente. También de esta forma se evitan acciones intrusivas a la maquinaria que como se ha comprobado muchas veces son generadoras de nuevos defectos. A modo de ejemplo, la figura 2.6 muestra la evolución del punto de falla y las técnicas tradicionales para el monitoreo de condición para un fallo de rodamientos.

Si bien muchos de los planes de mantenimiento basados en la condición de equipos se limitan a la vigilancia de sensores y sistemas expertos, los datos de proceso y otros datos históricos de mantenimiento, no se utilizan en forma integrada para la obtención de información más precisa y valiosa. Muchas veces por la dificultad de integrar datos con relaciones no siempre evidentes aún para el profesional medianamente calificado. Otro problema de los sistemas de diagnóstico, que desde hace años existen en mantenimiento, y son construidos en base a reglas aportadas de la experiencia de expertos, es que trabajan bien cuando la cantidad de datos es relativamente pequeña.

El Mantenimiento basado en la condición y la adquisición de datos

Con el avance del tiempo y los desarrollos tecnológicos, se abaratan los costos de los equipos de monitoreo, medida, control y almacenamiento de datos. El monitoreo de condición de los equipos se lleva a cabo hoy en día por diversos sistemas dedicados de distintos proveedores, algunos *online* y otros *offline*, lamentablemente no siempre bien integrados. Imágenes termográficas, niveles de vibraciones, emisiones acústicas, son algunos ejemplos de ellos (Fig.2.7).

Technologies	Applications	Pumps	Electric Motors	Diesel Generators	Condensers	Heavy Equipment/ Cranes	Circuit Breakers	Valves	Heat Exchangers	Electrical Systems	Transformers	Tanks, Piping
Vibration Monitoring/Analysis		X	X	X		X						
Lubricant, Fuel Analysis		X	X	X		X					X	
Wear Particle Analysis		X	X	X		X						
Bearing, Temperature/Analysis		X	X	X		X						
Performance Monitoring		X	X	X	X				X		X	
Ultrasonic Noise Detection		X	X	X	X			X	X		X	
Ultrasonic Flow		X			X			X	X			
Infrared Thermography		X	X	X	X	X	X	X	X	X	X	
Non-destructive Testing (Thickness)					X				X			X
Visual Inspection		X	X	X	X	X	X	X	X	X	X	X
Insulation Resistance			X	X			X			X	X	
Motor Current Signature Analysis			X									
Motor Circuit Analysis			X				X			X		
Polarization Index			X	X						X		
Electrical Monitoring										X	X	

Fig. 2.7 Tecnología de aplicaciones de mantenimiento predictivo.

Como fue mencionado anteriormente, hoy en día las plantas industriales están sumamente automatizadas, con *DCSs* (Sistemas de Control Distribuidos) y sistemas *SCADA* (*Supervisory Control And Data Acquisition*) de diversos proveedores como Honeywell, ABB, Foxboro, Siemens; con exigencias crecientes sobre mantenimiento en cuanto a confiabilidad,

disponibilidad de equipos, seguridad, medio ambiente y calidad de producto (Fig. 2.8). El número de activos físicos, complejidad, y grado de especialización requerida del técnico de mantenimiento y operación se encuentra en continuo aumento. Con su especialización aumenta el costo de la mano de obra y en consecuencia se reducen los recursos.

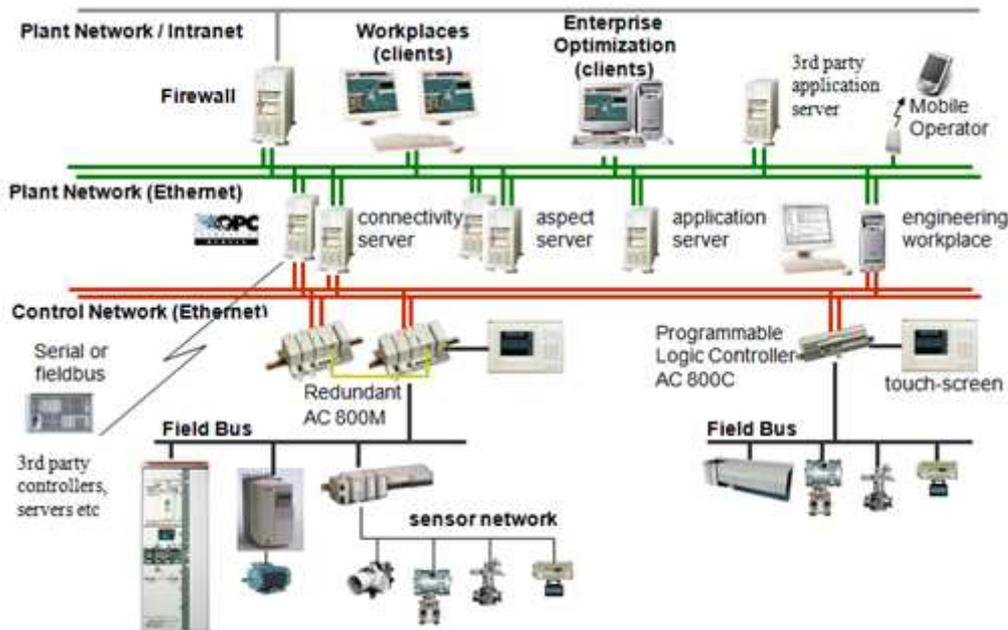


Fig. 2.8 ABB Industrial IT (DCS System)

En este contexto, el profesional de mantenimiento debe establecer políticas y planes de mantenimiento que sean efectivos, pero también eficientes. Los datos están siendo producidos y se encuentran disponibles. No así las habilidades, herramientas y destrezas necesarias para poder analizarlos. Se está tornando cada vez más común que luego de manifestada la falla, se analizan los datos, y como "Crónica de una muerte anunciada" se hallan registros que de haber sido analizados en tiempo y forma, la podrían haber predicho.

Un parada bien planificada en mantenimiento permite una optimización en la utilización y coordinación de los recursos (incluido repuestos), un mejor aprovechamiento del tiempo, menos pérdidas por costos de oportunidad de producción, menores probabilidades de accidente, y en resumen un mejor servicio al cliente (entendiendo a Producción como el cliente directo de mantenimiento).

2.3 Un problema de la vida real: “DATA SENSOR CHAOS”

A modo de ejemplo y con el objetivo de mostrar claramente el problema, se presenta un caso real reciente sucedido en una moderna planta de celulosa. A pesar de tener todos los datos disponibles para ello, no fue posible detectar a tiempo una falla en el equipo de mando de un motor. Esto trajo como consecuencia un desbalance de planta por disminución de chips de entrada y una pérdida de producción que se pudo haber evitado.

Para entender las cantidades de variables que interactúan en cualquier proceso industrial la (Fig. 2.9) muestra en forma simplificada una pantalla de monitoreo y control del patio de madera (*WoodHandling*) de una fábrica de celulosa. Se puede apreciar los diferentes actores y medidas (m³, velocidades, flujos de chips, etc.) más relevantes para el proceso, desde la entrada de madera en forma de troncos, hasta la salida en forma de chips al Digestor donde comienza la cocción para la producción final de pulpa de celulosa.

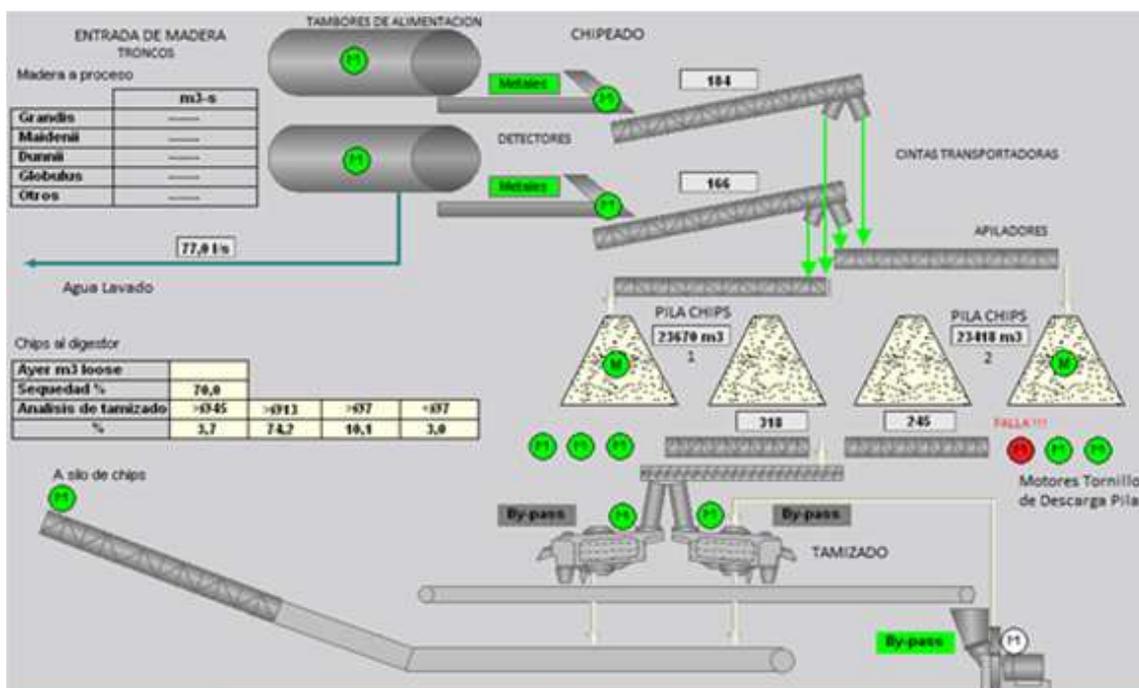


Fig. 2.9 Pantalla de control y monitoreo Patio de madera, (HMI).

El motor representado en color rojo en la figura anterior acciona uno de los tres motores de descarga del tornillo de la pila 2 a la cinta de tamizado. Este se detiene el 3/4/2012 a las 5 am por alta temperatura del equipo electrónico que controla la velocidad del motor (*VSD – Variable Speed Drive*). Se procede a inspeccionar y se encuentra placa electrónica de semiconductores *IGBT* dañada. Se decide cambiar el equipo y se libera para producción, quedando la función completamente recuperada 5 horas después.

Solucionado el inconveniente, posterior al problema se procede con el análisis de causa raíz, una herramienta fundamental del mantenimiento que tiene como fin encontrar la causa inicial que desencadenó en este caso la rotura de la placa electrónica de potencia del VSD, y que tiene como objetivo prevenir similares futuras fallas. Usando el *Plant Information Management Systems (PIMS)* de la fábrica, se obtienen las tendencias de las variables más significativas al problema; y para este caso se seleccionaron:

- Velocidad, temperatura , potencia y torque del motor de Descarga Tornillo Pila 2
- Producción de chips de la pila 2 - m3/t (#6)
- Temperatura del disipador (°C) del VSD del motor del tornillo descarga A (#1)
- Temperatura de sala eléctrica (no graficada)

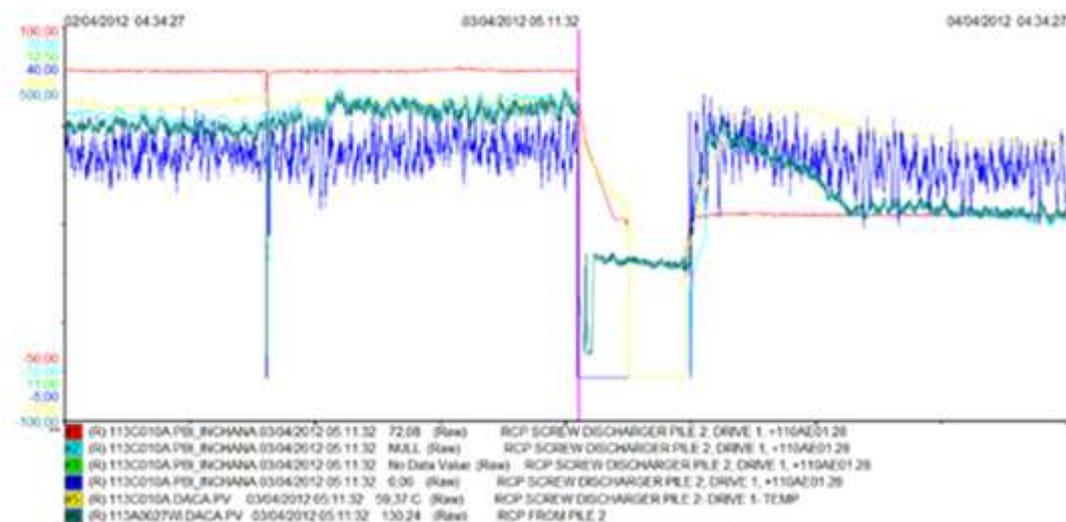


Fig. 2.10 Gráfico de tendencias variables de proceso (*PIMS*)

En la Fig. 2.10 se pueden ver las tendencias registradas de las variables anteriormente mencionadas, desde un mes previo y hasta un día después de solucionada la falla. El periodo entre las 5:11 am y 10:15 del 03/04/2012 representa el periodo de falla. Lo primero que llamó la atención fue que no se notaron cambios significantes en ninguna de las variables graficadas un mes atrás al momento de la falla, sobretodo la que muestra la temperatura del VSD (#1 – Línea roja). Sin embargo, una vez solucionado el problema y en situación de régimen un día después, dicha temperatura se redujo a casi la mitad.

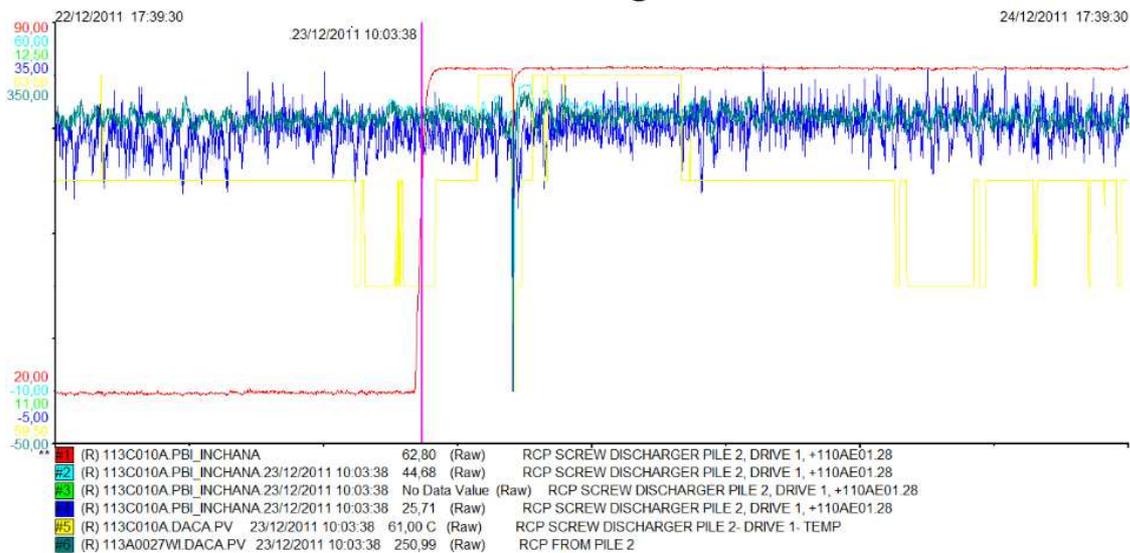


Fig. 2.11 – Tendencias PIMS 4 meses atrás

Lo anterior llevó a que se realizara una consulta en el *PIMS* de un período anterior más prolongado (más de 3 meses) de las mismas variables (Fig. 2.11) y se pudo apreciar con sorpresa que el 23/12/2011 a las 10:00 am se produce una subida de 20 a 80 °C sin ningún cambio operativo del proceso que lo justificara (Trend - #1), como una mayor velocidad del motor debido a una mayor producción. Evidentemente esta fue la fecha en la cual se rompió el ventilador de refrigeración del *VSD*.

Es así que el equipo y ante la sorpresa de todos, estuvo más de 3 meses con una falla de refrigeración sin dar alerta y sin que nadie pudiese advertir el problema. Para esto evidentemente ayudó que el equipo se encuentra sobredimensionado para la aplicación, trabajando a menos de la tercera parte de su capacidad nominal, y que el aire acondicionado de la sala eléctrica donde se encuentra el equipo, ayudó a mantener la temperatura controlada de la sala a pesar de las altas temperaturas de Enero y Febrero. Se pudo comprobar además que durante ese periodo existieron otras paradas del sector por otros mantenimientos que pudieron haberse aprovechado para hacer la intervención correctiva si el problema se hubiese podido detectar oportunamente.

Cuantificando el costo de oportunidad de la falla (Pérdida teórica de la producción)

Para cuantificar la pérdida basta con ver cuál es el nivel de producción durante el periodo falla, y compararlo con el que hubiera tenido si hubiese continuado produciendo sin inconvenientes. Se multiplica esto por el margen de beneficio por tonelada vendida y se calcula el costo que tuvo la falla. Por simplicidad se hace el cálculo de la pérdida solamente con las ventas de una fábrica de 1 millón de Toneladas año. Esta fábrica produce 2900 toneladas de celulosa por día en valor medio, y 5 horas de producción equivalen a 625 TON producidas, a un costo de 825 USD/TON en abril del 2012, equivaldrían entonces a 515.625 USD en ventas.

¿Justifica entonces invertir en diagnóstico y evitar las fallas?

Evidentemente este representa un caso simple de un motor que acciona un tornillo, pero sirve para mostrar claramente como la cantidad de variables puede llegar a ser de cientos a miles, demasiadas para el ojo humano como para poder discernir desviaciones y detectar fallas incipientes. Esta planta tiene más de 3000 motores para que se tenga una idea. Esto es lo que D.S. Doan denomina como *“Data Sensor Chaos”* [12].

En resumen:

La información contenida en los datos está llegando demasiado tarde para proteger a los equipo de costosas fallas y lo que es peor de potenciales mortales daños. Se necesitan entonces nuevas técnicas o herramientas. Obtener datos ya no es un problema económico, las plantas disponen de ellos y se debe empezar a utilizarlos. Esta es la principal motivación de la presente tesis, buscar nuevas herramientas para mantenimiento y en particular investigar sobre el uso de técnicas de minería de datos para poder dar solución al problema planteado.

“There is an Urgent need for a new generation of computational theories and tools to assist humans in extracting useful information (Knowledge) from rapidly growing volumes of digital data” [13]



SECCIÓN 3 - MANTENIMIENTO INDUSTRIAL Y *DM*.

3.1 Valor agregado del *DM* al mantenimiento

En esta sección se considera la importancia de *DM* al mantenimiento. Se intentará responder a las siguientes preguntas:

¿Es importante la minería de datos en mantenimiento? ¿Agrega valor? ¿Es viable en los entornos industriales?

Antes de responder estas preguntas se delimitará formalmente el tipo de industrias al que se hace referencia: Fábricas de proceso continuo, altamente tecnificadas, capital intensivas, con costos de oportunidad asociados a paradas de producción sumamente altos (Ej: fábrica de pulpa y papel, otras plantas químicas, refinerías, estaciones generadoras). Donde la parada planificada de mantenimiento total de fábrica se limita generalmente a un periodo acotado a unas pocas semanas al año, y en consecuencia con campañas de producción prácticamente anuales. En este escenario, todas las herramientas de predicción y/o análisis que permiten realizar ensayos o diagnósticos no invasivos, es decir conocer o predecir el estado de salud de las máquinas sin parar la producción, tienen un alto valor agregado para Mantenimiento y por supuesto para producción. Por lo tanto, se justifica la inversión en estas herramientas sin mayores análisis generalmente. Fue visto por ejemplo en el punto 1.3, los costos asociados a una parada de 5 horas de producción en una fábrica de celulosa.

Si bien se puede extender su importancia a cualquier tipo de industria en general, son en este tipo de industrias donde dadas sus exigencias operacionales, confiabilidad y calidad de producto, se brindan las condiciones básicas para que un proceso de *DM* sea posible: La existencia de grandes volúmenes de datos y la conveniencia económica de analizarlos. La minería no es una herramienta económica por su necesidad de requerir de sistemas de adquisición, de almacenamiento y capacidades computacionales intensivas. Con esta delimitación inicial nos aseguramos que los datos existan o que sea económicamente viable obtenerlos. El problema entonces se centrará en cómo utilizarlos.

Las principales y más comunes bases de datos (BD) disponibles en estos entornos industriales son:

- *PIMS (Plant Information Management System)*
- *CMMS (Computer Maintenance Management System)*
- Sistemas expertos dedicados (Emisiones acústicas, vibraciones mecánicas, etc.)
- *SCADA (Supervisory Control And Data Acquisition)*

De esta manera, si se logra a través de la minería de datos, obtener información de forma

integrada de todas estas bases de datos, y predecir el estado de los equipos, brindando soporte para la toma de decisiones de una forma más eficiente, se habrá justificado su importancia y el valor agregado del *DM* al mantenimiento de activos industriales.

Algunas de las preguntas que se intentará responder son:

- a) ¿Cómo mejorar la eficiencia de los planes y rutas de mantenimiento?
- b) ¿Podremos utilizar los datos del proceso para crear sistemas expertos más sofisticados?
- c) ¿Existe correlación entre fallas, actividades de mantenimiento, u otras condiciones de proceso?
- d) ¿Cuáles son los atributos o variables más significativas asociadas a los patrones de fallas de los equipos?

Metodología

Se mostrará el valor agregado del *DM* al mantenimiento industrial, a través de la presentación de algunos problemas planteados en mantenimiento y cómo mediante la aplicación de herramientas de minería podríamos obtener información útil para poder solucionarlos. Se buscó incluir ejemplos que incluyeran varias técnicas de *DM* diferentes, sobre equipos de uso común y que no sean particulares de un tipo de proceso en especial. De esta forma se logra darle una mayor generalidad a las conclusiones obtenidas.

Otro camino podría haber sido desarrollar un proceso de *DM* totalmente nuevo, comenzando por identificar una oportunidad de negocio en mantenimiento. Sin embargo esto requiere de una etapa crítica, por lo general mano de obra intensiva de transformación de los datos para prepararlos para un proceso de minería y muchos de los procesos fracasan, o no se obtienen buenos resultados debido justamente a la mala calidad de los datos. Para sortear esta dificultad se decidió por la metodología del párrafo anterior.

¿Cómo nos puede ayudar el *DM* en mantenimiento?

a) Identificando los modos de falla de los equipos (*Classification & Clustering*)

Identificar los modos de falla es una de las tareas esenciales del mantenimiento. Tiene como objetivo encontrar un patrón, un modo común y/o identificar qué atributos son significativos de que un equipo falle. Con esta información se puede hacer una utilización más eficiente de los recursos, atacando el origen de los problemas, y lo más importante, prevenir futuras fallas evitando que se repitan.

En términos de minería de datos esta búsqueda de modos de falla, puede traducirse en una tarea de clasificación (reglas de clasificación ó arboles de decisión) o segmentación (*clustering*).

La siguiente tabla muestra a modo de ejemplo algunas de las tantas y a veces cientos de variables o atributos que puede tener un equipo industrial, en este caso se ha seleccionado un

motor eléctrico presente en toda industria.

Objeto : Motor Eléctrico			
Origen Datos	Variable name	Description	type
CMMS	ID	Código de identificación	Number
CMMS	Departamento	Ubicación del equipo en la planta (área)	Set
CMMS	Marca	Fabricante	Set
CMMS	Modelo	Frame	Set
CMMS	Potencia	KW nominales	Number
CMMS	Aplicación	(bomba, tornillo, rodillo, transportador, agitador, etc)	Set
PIMS	Mando	(Variador de velocidad, Partida directa)	Set
PROJECT DATA BASE	Ubicación	(Exterior (bajo lluvia) , Interior)	Set
PROJECT DATA BASE	Refrigeración	(Ventilador trasero, Aire/aire, Aire/agua)	Set
CMMS	Rodamientos	(bolilla, rodillo, buges, cerámicos)	Set
PIMS	Criticidad	Factor de criticidad	Numerical
SCADA	CURRENT_AVGE	CORRIENTE MEDIA	Number
SCADA	Utilización	(Continua, intermitente, respaldo)	Set
SCADA	POWER_AVGE	Potencia Media	Number
SCADA	Start_hr	Arranques por hora	Numerical
CMMS	Acción	Actividad de reparación	Text
PIMS	Turno	Turno en que se produce la falla	Numerical
.....	otras

Si bien es cierto que a partir de los atributos se pueden deducir muchas relaciones por métodos deductivos tradicionales (dependiendo de las capacidades de los especialistas), esto se vuelve cada vez más impracticable o imposible cuando el número de variables aumenta. Por otro lado los sistemas expertos industriales, suelen mirar solo algunas de ellas, una parte de la película, generalmente pertenecientes al mismo origen de datos.

La minería en cambio permite integrar de forma "simple" datos de diferentes orígenes, sin la necesidad de establecer relaciones, o de ser un experto en el dominio de los equipos; y obtener en consecuencia, de forma casi automática, relaciones o patrones comunes de la fallas. Es tarea posterior de la persona que los analiza, dar el peso necesario a los mismos, discriminando trivialidades, incoherencias y por supuesto lo más importante, filtrar y reconocer la información útil.

Obtención de Información con métodos de *DM*:

- Supervisado : Clasificación

Se aplica un método supervisado de clasificación (C5.0) a los datos de objeto Motor y se define

como atributo de clase o etiqueta al atributo *Acción*, la cual contiene Texto (previamente codificado) de la reparación realizada. Más allá de la simplicidad del ejemplo elegido especialmente con fines didácticos, sería de esperar obtener una regla de clasificación del tipo:

IF Ubicación = "Exterior" & Utilización = "Respaldo" THEN Acción = "Motor Húmedo"

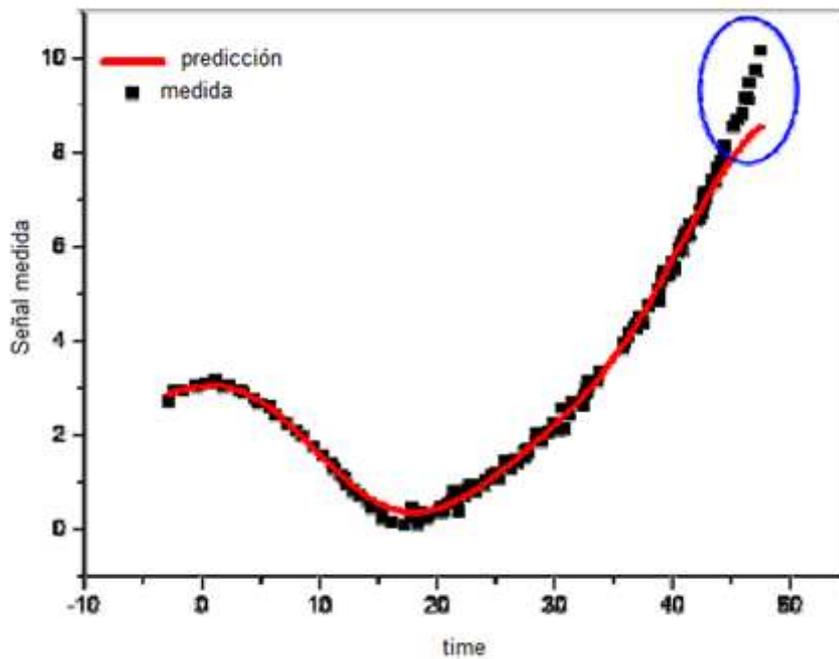
Aún este ejemplo que parece no aportar mucha información al especialista de mantenimiento, ya que si el motor es exterior tiene mayor probabilidad de humedecerse, y más aún si permanece mucho tiempo parado, como generalmente es el caso de los motores de respaldo; sin embargo, puede significar una falla sistemática de mantenimiento en la instalación y/o selección de las juntas de sellos de las tapas de los motores por ejemplo.

- No Supervisado: *Clustering*

Si no identificamos ningún atributo con característica de etiqueta o clase, es decir no disponemos de ninguna información o sospecha de por dónde empezar a buscar, por medio de un algoritmo de *clustering* (Ej: Kmeans), método no Supervisado de *DM*, se puede segmentar los datos y obtener conglomerados o *clusters* que pueden llegar a representar características interesantes. De esta forma, se podrían llegar a identificar grupos de datos (en este ejemplo motores) que comparten ciertos valores de atributos, muy similares entre si, y muy diferentes con los de los otros conglomerados ó *clusters*.

b) Prediciendo el mal funcionamiento de un máquina (Time Series [14])

Poder predecir o estimar el mal funcionamiento de un equipo basado en los datos obtenidos de los sensores instalados en las máquinas puede ahorrar mucho dinero debido a la posibilidad de prevenir futuros daños. La materia prima fundamental que nos proveen los sensores conectados a los equipos, como señales acústicas, vibración, temperatura, presión o tensión, están constituidas por series de valores en el tiempo denominadas *time series*. La minería de datos posee poderosas técnicas de regresión disponibles para la predicción de valores de este tipo. La Fig. 3.1 nos muestra una serie de valores medidos por un sensor, y conjuntamente una predicción basada en el método denominado *Support Vector Machine (SVM)*. Evaluando el desvío entre la predicción y el valor medido se puede predecir el mal estado del equipo.

Fig 3.1 Regresión de *time series*

¿Cómo el *DM* aprende y crea esta predicción?

El método se denomina ventaneo (*Windowing*), y consiste en usar valores anteriores de la señal medida en un intervalo de ventana de muestras, para predecir las señales futuras luego de un intervalo denominado horizonte (*Horizon*). Moviendo la ventana paso a paso y manteniendo el horizonte constante es que se crea el conjunto de datos para predecir el futuro (Fig. 3.2)

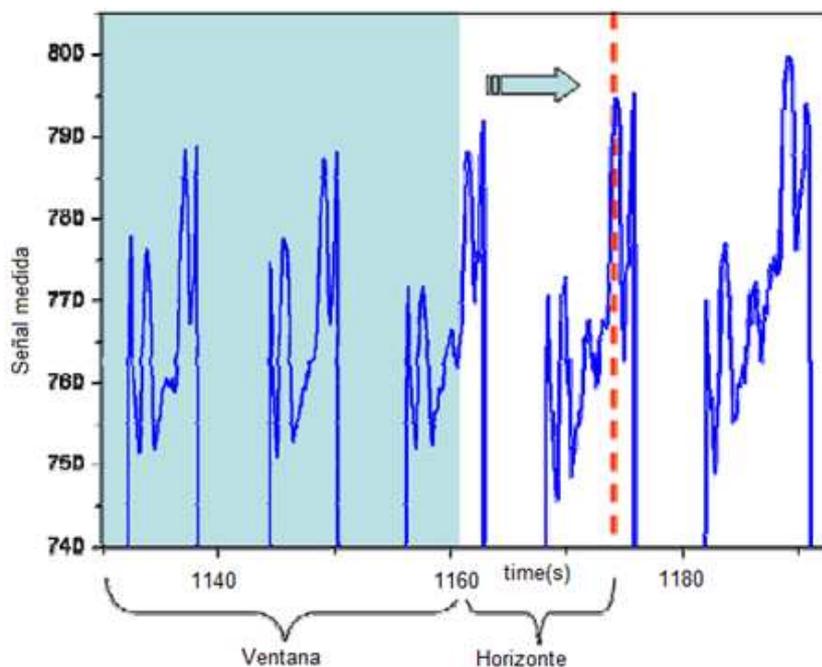


Fig. 3.2 Ventaneo

La Fig. 3.3 muestra un ejemplo de aplicación que explica las ventajas que la predicción tiene relacionado con el monitoreo de condición: Sensor que mide la vibración de una máquina y detecta un incremento. Un sistema de monitoreo de condición por sí solo no puede predecir el futuro, haciendo imposible juzgar si el incremento va a ir valores críticos y cuándo. Este inconveniente puede ser resuelto por una predicción realizada por *time series* usando la técnica de *DM* explicada anteriormente. Según ésta es claramente visible que la vibración se va a estabilizar y nada va a suceder.

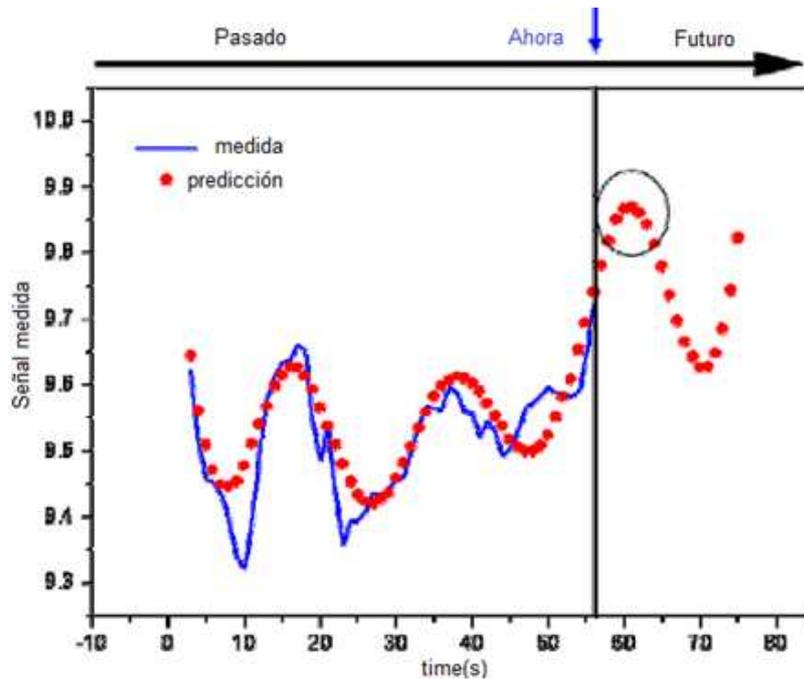


Fig. 3.3 Predicción de *Time Series*.

El esquema de la Fig. 3.4 muestra cómo se puede realizar una implementación de este sistema en una planta industrial. Los sensores instalados en las máquinas monitorean el proceso. Estos pueden medir diferentes clases de señales (temperatura, vibración, fuerza, tensión, humedad, emisión acústica, etc.). Una interfase controla la adquisición de datos y almacena la *time series* en la base de datos (DB). Estos son los datos de entrada para el software de *DM*, que realiza la predicción y muestra los resultados para el usuario.

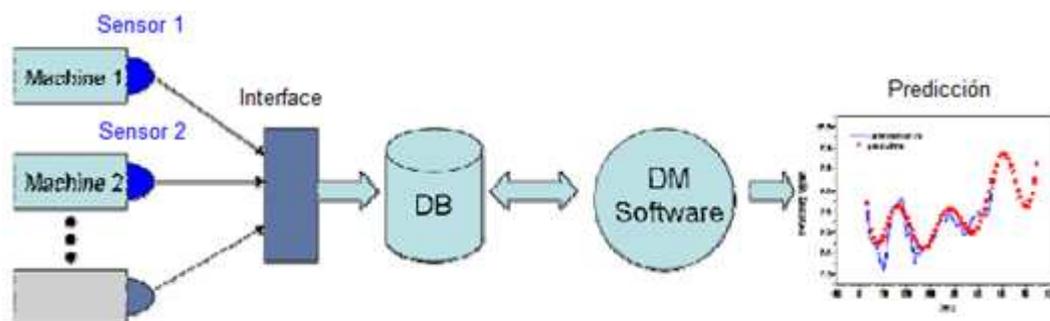


Fig. 3.4 Arquitectura del sistema de *DM* en una planta industrial

- c) Aportando información para la toma de decisiones en mantenimiento de forma eficiente. (Clasificación, *Clustering* y técnicas de visualización)

Poder predecir el estado de condición de los equipos y planificar las tareas de mantenimiento necesarias en forma eficiente y efectiva, es fundamental para mantenimiento. Generalmente el tiempo y mano de obra son recursos escasos. El siguiente caso de estudio muestra como la minería de datos puede ayudar a establecer el estado de condición y dar soporte a la toma de decisiones en mantenimiento. Aquí se presenta el problema y las soluciones planteadas por el *DM*, se trató de evitar profundizar en aspectos técnicos de los equipos o de la minería. Por mayor información se sugiere consultar la referencia [15]

Problemática:

Los interruptores de potencia son clásicos equipos que necesitan estar la mayor parte de tiempo cerrados, "sin actuar", para asegurar la alimentación de energía de las cargas que alimentan. Ellos deben asegurar que se transmita energía aguas abajo pero deben ser capaces de abrir o interrumpir el circuito cuando se produzca una falta para poder despejarla. En estas circunstancias deben actuar en forma rápida y segura. Probar este tipo de equipos es sumamente complicado generalmente, ya que hay que simular la falta y cortar el suministro. Más complicado aún es cuando se trata de empresas suministradoras de energía, compitiendo en mercados fuertemente competitivos por la confiabilidad y calidad de los servicios a sus clientes. El mantenimiento clásico de estos equipos consiste en inspecciones profundas generalmente anuales, sacando al interruptor fuera de servicio y realizando la inspección. Si al abrir se encuentran componentes en mal estado se habrá justificado el tiempo y el trabajo, el problema es cuando todo está bien y el mantenimiento no aportó nada, o lo que es peor puede ser introducida una falla por las tareas realizadas (Ej. un mal armado).

Objetivo:

Este caso de estudio propone una metodología como soporte a la toma de decisiones para determinar que interruptores es necesario inspeccionar a partir del análisis de los datos obtenidos durante los ciclos de apertura de los mismos. En particular se analizará las características de la forma de onda de corriente que pasa por la bobina de apertura del interruptor. Usando el conocimiento experto disponible, en combinación con un estructurado proceso de minería de datos, y los tiempos asociados con las características que representan cada etapa de la operación del interruptor, se llega a determinar el estado de condición del interruptor.

Preparación de los Datos:

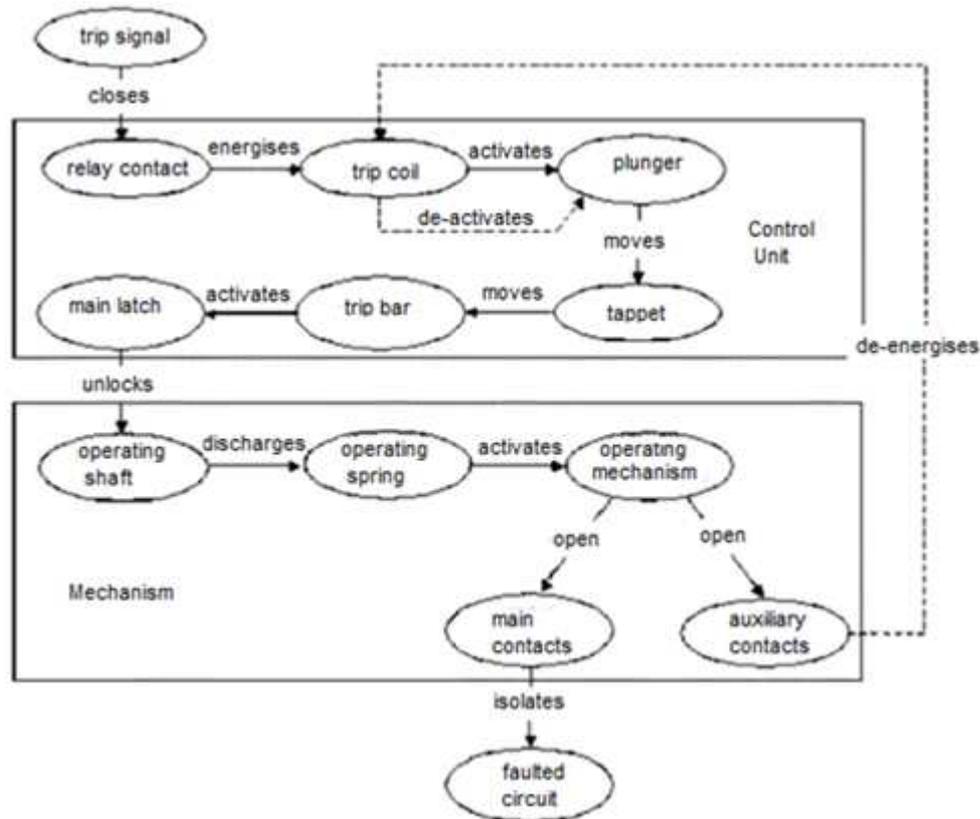


Fig. 3.5 Estados durante la operación de un interruptor

La Fig. 3.5 representa el diagrama de estados de la operación de un interruptor de potencia. Este busca explicar en forma de diagrama de flujo las diferentes etapas de un ciclo de apertura de un interruptor.

La Fig. 3.6 presenta la forma de onda de la corriente por la bobina de apertura en cada etapa, desde que el *Trip-Switch* es accionado hasta que el interruptor abre totalmente. Esta forma de onda es registrada con un *Datalogger*, de uso común en la industria y fueron seleccionados 5 atributos para determinar el vector de estado de condición del interruptor [16]:

- *Latch*: Tiempo requerido para que la bobina de disparo libere el pestillo para iniciar mecanismo.
- *Buffer*: Tiempo en el que el pestillo se desprende y el mecanismo actuador comienza a moverse.
- *Mcon*: Tiempo en que los contactos principales se separan.
- *Acon*: Tiempo en el que el contacto auxiliar comienza a abrirse y la corriente por la bobina reducirse a cero.
- *End*: Tiempo en el que la corriente de la bobina se extingue.

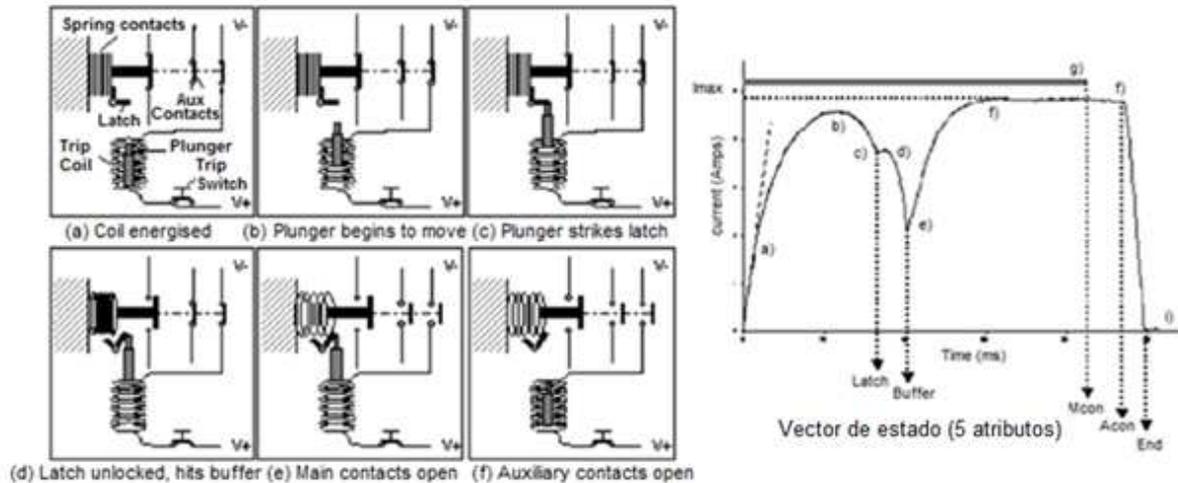


Fig. 3.6 Diferentes etapas en la apertura de un interruptor y forma de la corriente a través de la bobina de apertura.

Con estos atributos, y después de obtener el vector de condición para toda la población de los interruptores que se desean analizar, se obtiene el conjunto de datos sobre el cual se basa el presente estudio. El proceso de *DM* que se lleva a cabo se muestra en la fig. 3.7.

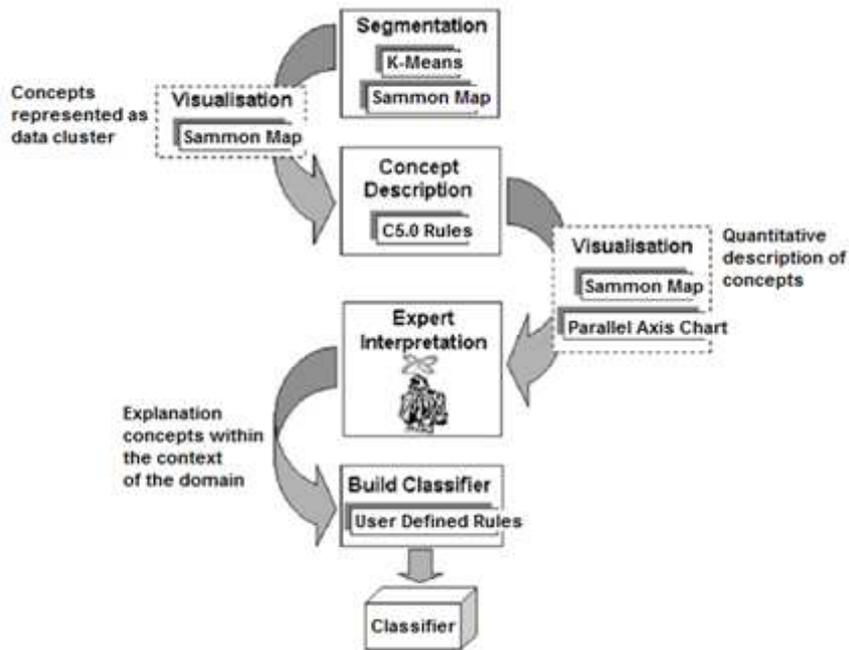


Fig. 3.7 Proceso de *DM* para el análisis de los datos de la corriente por la bobina de apertura y desarrollo del clasificador.

Sin querer entrar en detalles sobre los algoritmos se dirá que *Sammon Map* es un método de mapeo que permite representar un espacio de dimensión “n” en uno de menor dimensión, con ciertas características que minimizan los errores producidos de esta proyección. En particular en este caso, interesa poder mapear un vector de dimensión 5 en uno de dimensión 2, que

permita interpretar visualmente los resultados. La figura 3.8 (a) presenta los puntos de proyección por *Sammon Map* de los datos-vectores de la corriente de apertura de los interruptores. En él se pueden apreciar 4 diferentes *clusters*, este número es utilizado para inicializar el número de *cluster* que necesita el algoritmo de *Kmeans*. La Fig. 3.8 (b) presenta el resultado de los *clusters* encontrados *Kmeans*, sobre el mismo gráfico del mapeo de *Sammon*, y donde los puntos con diferentes formas pertenecen a distintos *clusters*. Es evidente la similitud de los resultados encontrados por ambos métodos con respecto al comportamiento de aglomeración (clustering) de los datos.

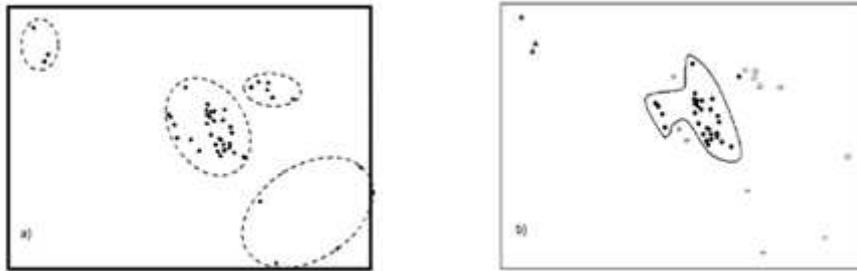


Fig. 3.8 (a) *Sammon Map* de los vectores de estado de los interruptores y Fig.3.8 (b) Resultado *K-means clustering* superpuesto con *Sammon Map*.

El método de *K-Means* identifica los *clusters* representativos de los estados de condición de los interruptores específicamente y no se basa en la identificación visual de los *clusters* como es el caso con el enfoque de *Sammon Map*.

¿Pero cómo relacionamos los *clusters* y el estado de condición de los Interruptores?

Las Fig. 3.9 a) y b) nos muestran las relaciones existentes entre los diferentes resultados. Los gráficos de coordenadas paralelas son otra forma interesante de representar una serie temporal (*time serie*). Se puede apreciar la correspondencia entre gráficos de coordenadas paralelas y los puntos de los *clusters*.

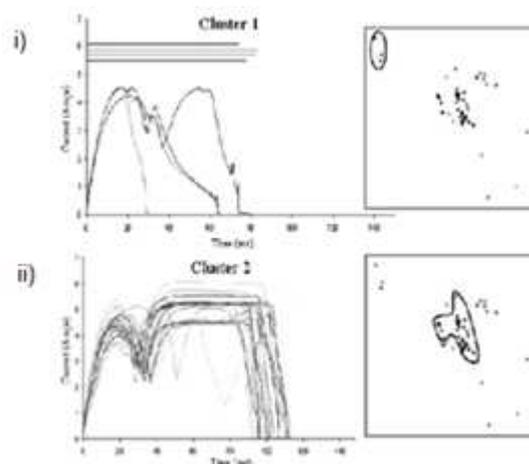


Fig. 3.9 a) *Clustering*: formas de corriente, bobina de apertura

En el gráfico 3.9 (b) también se pueden apreciar las gráficas de distribución relativa de los valores de los 5 atributos del vector de estado.

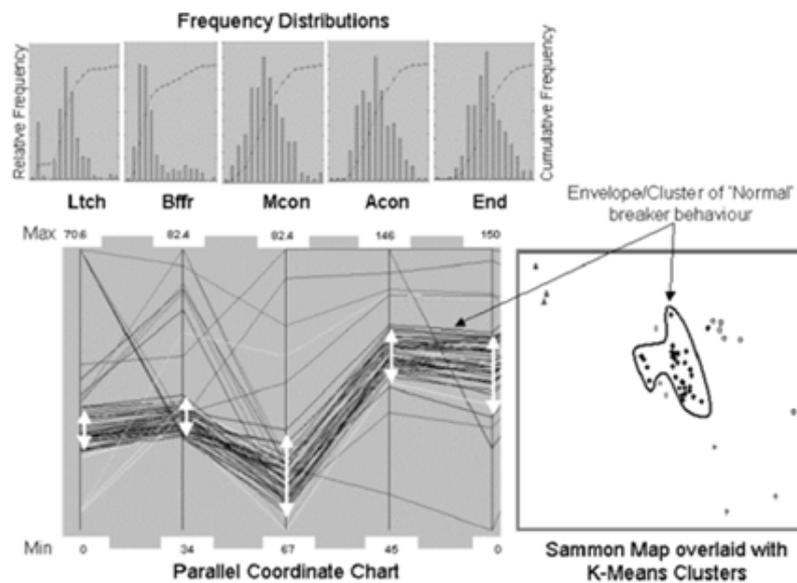


Fig. 3.9 (b) Gráficos de coordenadas paralelas describen las propiedades de los vectores de estado de condición.

En la Fig. 3.9 (a) i), se puede apreciar algunas formas de onda de algunos interruptores pertenecientes a un mismo *cluster*, que difieren bastante del resto. El método supone que la mayoría de los interruptores se encuentran en buen estado por lo cual este *cluster* y todos sus integrantes se podrían clasificar como defectuoso y necesitaría mantenimiento. De la misma forma, se puede analizar cada uno de los *clusters* y cada una de las formas de onda. Para el experto la forma de onda puede representar un problema grave o simplemente una alarma, existe una correspondencia entre las formas de onda y el tipo de problema que la causa. Por ejemplo, una corriente de bobina que decae a cero rápidamente puede significar tan solo una alta resistencia de contacto en el interruptor que la acciona apenas ésta empieza a abrir. Con estas observaciones se puede agregar un atributo de clase a los datos y clasificarlos por ejemplo mediante un algoritmo C5.0.

Si por otro lado, no se dispone de los conocimientos técnicos o pericia, para diagnosticar cual forma de onda puede ser representativa de un problema grave que necesita ser atendido, otra forma de definir un clasificador por ejemplo con tres estados: ROJO, AMARILLO, VERDE, es utilizar la función de distribución de alguno de sus atributos y definir, con la ayuda de los expertos, que porcentaje de los interruptores se va a considerar por hipótesis en cada una de estas categorías. Por ejemplo supongamos que para una población de interruptores de un determinado modelo, el experto estableció que según el tiempo de apertura de los contactos auxiliares los mismos se clasifican de la siguiente forma:

- 70% como normal (VERDE)
- 20% con defecto no urgente (AMARILLO)

- 10% con falla (ROJO)

Con esta hipótesis, y con la ayuda del gráfico de distribución de frecuencia del atributo tiempo de operación de contactos auxiliares, se obtienen los valores de umbrales para programar el clasificador, Fig. 3.10.

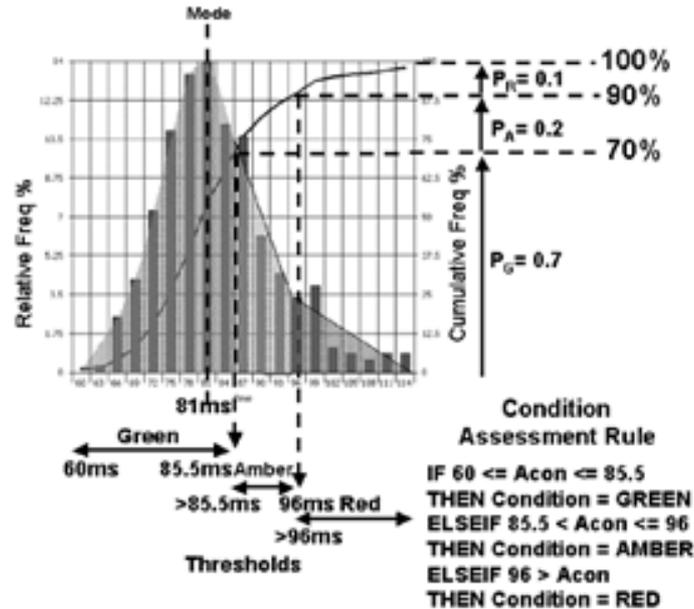


Fig. 3.10. Distribución de frecuencia (función de densidad de probabilidad) para el atributo tiempo de operación de los contactos auxiliares.

El resultado final de este estudio fue la construcción de un modelo clasificador y la implementación de un sistema de soporte a la toma de decisiones (*DSS*).

¿Cuáles fueron los aportes de este clasificador/*DSS* a la gestión de mantenimiento?

- Más rápida, automatizada y consistente evaluación del estado de condición de los interruptores (*Circuit Breaker- CB*).
- Detección más temprana e incipiente de fallas.
- Relación entre *CB* / fabricante y tipo de fallas.
- Archivar información confiable y representativa, detallando la condición del activo.
- Aportar información para el diagnóstico, y para la toma de decisiones.
- Evitar el mantenimiento innecesario y fallos consiguientes.
- Reducir costos de mantenimiento.

En resumen, las preguntas planteadas al inicio de esta sección se responden con el desarrollo de los ejemplos anteriores y sus resultados. Podemos inferir entonces, que la Minería de Datos es importante para el mantenimiento industrial, agrega valor, y complementa las herramientas y técnicas utilizadas tradicionalmente.

3.2 Breve sondeo sobre el estado actual de aplicación de estas herramientas en mantenimiento.

Comprobada las potencialidades del *DM* en mantenimiento, surge la duda si la apreciación, sobre el bajo nivel de utilización de estas herramientas en mantenimiento tiene justificación, o es simplemente una visión personal sesgada por la experiencia profesional. Con el objetivo de confirmar preliminarmente dicha apreciación y sin querer tener la rigurosidad de una encuesta exploratoria sobre el nivel de aceptación que tienen estas herramientas en el ámbito industrial, se realiza una encuesta interna en industrias de Uruguay y un sondeo en un grupo de Mantenimiento de la red social *Linkedin*.

Encuesta interna en industrias de Uruguay

Cuestionario: ¿Utiliza algún tipo de herramientas de aprendizaje de minería de datos (exceptuando estadística clásica) para la obtención de información útil a la gestión del mantenimiento?

- SI
- NO

Población: Profesionales de Mantenimiento pertenecientes a 12 Industrias de primera línea en Uruguay, de diferentes tipos: Químicas (pulpa y papel, pinturas), Arroceras, Azucarera, Eléctrica, Cementeras, Agrícola, Siderúrgica, y otras industrias manufactureras.

Procedimiento: Encuesta telefónica

Observación: Por las características de escala del mercado uruguayo, es relativamente simple a partir del *networking* profesional y actividad universitaria, poder obtener información en forma confiable y de primera mano de responsables y/o profesionales encargados sobre las herramientas que utilizan en mantenimiento en sus industrias.

Resultado: En el 0% de las 12 industrias se utilizan métodos de aprendizaje supervisado o no supervisados de *DM* en mantenimiento. A pesar de coincidir que es creciente el número de datos disponibles, se limitan a la utilización de estadística clásica para analizarlos. El 10% dice conocer estas herramientas relacionadas con técnicas de *CRM* y/o *Marketing*.

Sondeo Grupo LinkedIn

En el link detallado más abajo se podrá ver el sondeo y el debate iniciado en la red social *LinkedIn* en su totalidad:

copiman – comité panamericano de ingeniería de mantenimiento.

http://www.linkedin.com/groups?home=&gid=2108781&trk=anet_ug_hm&goback=%2Egmp_2108781

Pregunta inicial planteada: ¿Utiliza técnicas de *DM* para la gestión del Mantenimiento?



Observaciones: Uno de los votantes por SI solamente aplica métodos estadísticos

Conclusiones:

De los resultados obtenidos, y de los comentarios realizados en el debate por otros participantes del grupo se confirma que actualmente no es muy utilizado el *DM* por la ingeniería del mantenimiento. Fue creado otro sondeo en un grupo de habla inglesa de *LinkedIn*, denominado *Condition Monitoring & Electrical Diagnostic Testing* obteniendo resultados similares.

Nota: En el 2009 sólo el 8% de los estudios en el campo de aplicación de *DM* en manufactura eran en relacionadas con Mantenimiento. [17]

SECCIÓN 4 - DIFICULTADES, REQUERIMIENTOS, ¿QUÉ PUEDE SALIR MAL EN UN PROCESO DE *DM*?

En la sección anterior se vio la importancia y la poca utilización aparente del *DM* en mantenimiento. Surgen entonces las siguientes preguntas:

¿Por qué? ¿Cuáles son las dificultades que presenta un proceso de minería de datos en los ambientes industriales para ser implementado?

A continuación se presentan algunos factores que pueden explicarlo. Fueron seleccionados en base a la experiencia profesional, y a los comentarios vertidos por otros profesionales en diferentes debates sobre este tema (Ejemplo: ver sondeo mencionado en Sección. 3.2)

1. Falta de personal calificado y/o dedicado.

En el ambiente industrial, es frecuente la dificultad de poder contar con personal capaz de llevar adelante estas técnicas con resultados que se consideren útiles para fines de toma de decisiones. Si bien no requiere ser un *Phd.* en matemáticas para poder aplicar herramientas de *DM*, sí es necesario contar con ciertas aptitudes estadísticas y computacionales.

Por otro lado existe una gran inercia aún hoy en día en muchas empresas (heredado de la primera generación de mantenimiento), a continuar trabajando como bomberos, es decir “apagando incendios”. Donde el personal que está involucrado en la rutina diaria de atender las emergencias, es la misma que tiene que dedicarse a estudiar la forma de solucionarlos definitivamente. Por un lado entonces tenemos falta de personal calificado, y por otro personal calificado abocado a atender la rutina diaria correctiva (ó preventiva), sin tiempo específico para tareas exploratorias de minería de datos.

2. Mala calidad de datos.

Justamente por esta “poca necesidad” de utilizar los datos, es que no muchos se preocupan de la calidad de los mismos al realimentar el sistema. Estamos “llenos” de datos pero poder obtener información útil de los mismos puede llegar a ser imposible o requerir una tarea de limpieza, pre-procesamiento y transformación tan intensa que desalienta a cualquier minero. Un problema típico es el registro de la órdenes de trabajo, en ellos el técnico de mantenimiento realimenta el sistema con los resultados o acciones tomadas en cada actividad. La libertad de los campos con textos libres y la multiplicidad de criterios de los usuarios para realizar los reportes, hacen que trabajos similares puedan tener registros totalmente diferentes, y en consecuencia hacer imposible su clasificación posterior. Si a esto le sumamos campos que no se llenan, o que se completan con datos por defecto, no coherentes o incompatibles con los ingresados, entonces tenemos un caos. Esto se representa con la conocida frase “*Garbage*

input, garbage output” en el campo de la tecnología de la información. Es decir si entra basura, sale basura. Aún un proceso de minería que no sea exitoso a causa de la mala calidad de los datos también suma, ya que por lo menos puso en evidencia esta situación para poder mejorarla.

3. *ROI* difíciles de calcular e inmediatez de resultados gerenciales.

Los procesos de minería de datos enfocados en el negocio, no son de corto plazo, y a la vez son procesos cíclicos como se vio en el punto 1.3. Medir el retorno sobre la inversión en un ambiente industrial puede no ser una tarea fácil ya que generalmente son problemas de variables con entradas múltiples. Se estima que una reducción de costos del orden del 10% en mantenimiento tiene el mismo efecto que un incremento en ventas del orden del 40 %, pero aún así las empresas siempre son reticentes a invertir en cualquier sistema que no ofrezca resultados inmediatos. Esto no solo se aplica a las inversiones en *DM* por supuesto.

4. Desconocimiento. Zona de confort gerencial, mandos medios y expertos.

Los tradicionales métodos conducidos por hipótesis de la ingeniería Vs los métodos conducidos por los datos del *DM*, hacen salir de la zona de confort a muchos profesionales de mantenimiento. Para peor muchas personas tienden a descalificar lo que no comprenden, o bien creen que pudiera disminuir o amenazar su autoridad. Como ya fue visto, un proceso de *DM* necesita de la evaluación de un experto para la validación primaria de los resultados, por sí sola la minería no puede hacer nada, no sustituye a la inteligencia humana. Entonces si los expertos o personas de más experiencia no creen en ella entonces es imposible que la sinergia se produzca. En conclusión ni siquiera se inicia el proceso.

5. Malas experiencias. Resultados inútiles.

Evidentemente un proceso con un resultado erróneo ó sin aparente utilidad condiciona catastróficamente futuros intentos. Más aún si genera grandes pérdidas económicas por decisiones desacertadas, además de lo invertido en el proceso de *DM* por supuesto. Es bueno tener claro que un proceso de minería puede llegar a fallar por dos motivos [4]:

- Llegar a conclusiones que no son ciertas,
- Llegar a conclusiones que son ciertas, pero no útiles.

SECCIÓN 5 – CONCLUSIONES, CONSIDERACIONES Y RECOMENDACIONES FUTURAS.

El *DM* emerge como una herramienta de gran potencial, en un mantenimiento industrial inundado de datos pero ávido de conocimiento. Ayuda a predecir y diagnosticar el estado de condición de los equipos, brinda información útil para la toma de decisiones, y finalmente conocimiento para una gestión eficiente de los recursos.

Sin embargo no se debe esperar del *DM* más de lo que es, una herramienta complementaria a otras tantas existentes en mantenimiento. Como tantas otras no sustituye al experto, necesita de él para poder construir y validar los modelos o las reglas obtenidas. Tampoco se debe pensar que es una tarea fácil o automática, se requiere esfuerzo, método y constancia para llevar adelante un proceso de minería de datos, “no existen atajos para el conocimiento”.

Se ha mostrado la viabilidad de su aplicación, y es de esperar que poco a poco vaya ganando su espacio en mantenimiento industrial, como lo ha hecho ya en otras disciplinas (Genética, Meteorología, Control de calidad, Salud [18], Manufactura [19], etc.) No obstante, para ello se debe primero convencer a las gerencias del valor de las aplicaciones de tecnología de la información en general, y en particular del *DM*. Por ejemplo mediante pequeños proyectos, mostrar en resultados, algunos de los valores agregados anteriormente mencionados. En segundo lugar, se debe alimentar al sistema con datos de calidad, para lo cual es imprescindible involucrar a responsables y usuarios, tanto de producción como de mantenimiento, ya que sin convencimiento y compromiso no es posible obtener resultados. Si no convencemos a las personas que una buena calidad de datos es fundamental para poder obtener información útil en el futuro, pocos se preocuparán de ingresarlos en forma correcta. En tercer lugar debemos dar participación al personal de IT, como soporte fundamental para lograr arquitecturas que permitan una integración e intercambio de datos efectivos.

Por último, y a modo de anécdota, no hay que esperar que el *DM* resuelva problemas por sí solo, hay que tener cuidado con las conclusiones obtenidas: “El 20 por ciento de las personas muere a causa del tabaco. Por lo tanto, el 80 por ciento de las personas muere por no fumar. Así queda demostrado que no fumar es peor que fumar”.



BIBLIOGRAFÍA

- [1] Halkidi, M. (2000); *Quality assessment and Uncertainty Handling in Data Mining Process*. Recuperado en febrero 2012 de:

<http://www.edbt2000.uni-konstanz.de/phd-workshop/papers/Halkidi.pdf>
- [2] Crespoa, F; Weberb, R. (2004); *A methodology for dynamic data mining based on fuzzy clustering*. Recuperado en febrero 2012 de:

http://cabierta.uchile.cl/revista/31/mantenedor/sub/articulos_3.pdf
- [3] Berry, J.A; Linoff, G.S. (1997); *Data Mining Techniques: for marketing, sales and customer support*, (p 5), *Wiley Publisher*.
- [4] Berry, J.A; Linoff, G.S. (2011); *Data Mining Techniques for Marketing, sales and Customer Relationship Management, Third Edition*, *Wiley Publisher*.
- [5] Kantardzic, M. (2002); *Data Mining: Concepts, Models, Methods, and Algorithms*, *Wiley- IEEE Press*.
- [6] Molina L.C; (2001); *Data Mining: torturando a los datos hasta que confiesen*. Recuperado en febrero 2012 de:

<http://www.uoc.edu/web/esp/art/uoc/molina1102/molina1102.html>
- [7] Berson A; Smith S.J. (1997); *Data Warehousing, Data Mining & OLAP*, (p 341), *Mcgraw–Hill Publisher*.
- [8] *The Economist; A Golden Vein*, (2004). Recuperado en marzo 2012 de:

<http://www.economist.com/node/2724407>
- [9] Han, J; Kamber, M. (2006); *Data Mining-Concepts and Techniques*, *Elsevier Publisher*.
- [10] Moubray, J. (2001); *RCM II, Reliability-centered maintenance*, *Industrial Press Inc. Publisher*.
- [11] Mobley, K. (1990); *An introduction to predictive Maintenance*. *Elsevier Publisher*.

- [12] Doan, D.S., (2009); *Predictive Analytics in Equipment. Reliability Programs*. Recuperado en abril de 2012 de:

www.focus-nuclear.com/pdf/TP_doan.pdf
- [13] Fayyad, U. M; Shapiro, G. P; Smyth, P. (1996); *From Data Mining to Knowledge Discovery in Databases*, 0738-4602-1996, *AI Magazine* (pp 37–53).
- [14] Bergmann, A. (2010); *Data Mining for Manufacturing: Preventive Maintenance, Failure prediction, Quality Control*. Recuperado en abril de 2012 de:

<http://www.ndt.net/article/ndtnet/2010/bergmann.pdf>
- [15] *IEEE Transactions on Power Delivery*, (2007); *Providing decision support for the condition-based maintenance of circuit breakers through data mining of trip coil current signatures; Vol. 22, No. 1, (pp. 178-186)*.
- [16] Watson, R. (2000); *Diagnostic technique facilitates timing and condition monitoring of circuit breakers, presented at the Engineering Asset Management Conf., London, U.K.*
- [17] Choudhary, A.K., Harding, J.A., Tiwari M. K. (2009); *Data Mining in manufacturing*, (pp. 501-521), *Springer Publisher*.
- [18] Koh, H. C., Tan, G. (2005); *Data Mining Applications in Healthcare, Journal of Healthcare Information Management, Vol. 19, No 2*.
- [19] KUSIAK, A; (2006); *Data Mining for manufacturing and service, International Journal of Production Research; Vol. 44, Nos. 18–19, 15, September – 1 October 2006, 4175–4191*.

ANEXO - OTRAS APLICACIONES DE *DM* EN MANTENIMIENTO

- *Practical applications of data mining in plant monitoring and diagnostics.*

Strachan, S.M. and Stephen, B and McArthur, S.D.J. (2007) *Practical applications of data mining in plant monitoring and diagnostics. In: IEEE Power Engineering Society General Meeting 2007, 1900-01-01.*

<http://strathprints.strath.ac.uk/12871/>

- *Prediction of Status Patterns of Wind Turbines: A Data-Mining Approach.*

Andrew Kusiak and Anoop Verma, J. Sol. Energy Eng. -- February 2011 -- Volume 133, Department of Mechanical and Industrial Engineering, University of Iowa, 3131 Seamans Center, Iowa City, IA 52242-1527

http://ir.uiowa.edu/cee_pubs/464/

- *Data Mining Techniques to Assess the Condition of High Voltage Electrical Plant.*

Allan, David, Birtwhistle, David, Blackburn, Trevor, Groot, E, Galski, E, & McGrail, A (2002) *In Set of Papers Presented to the CIGRE 2002 Session No 39, 25 - 30 August 2002, Paris, France.*

http://www.gbook.yolasite.com/resources/15_107E.PDF

- *Diagnosis and prognosis of bearings using data mining and numerical visualization techniques.*

Blair, J.; Shirkhodaie, A. (2001) *Dept. of Mech. Eng., Tennessee State Univ., Nashville, TN.*

<http://ieeexplore.ieee.org/iel5/7335/19863/00918553.pdf?arnumber=918553>

- *Preventing Failures by Mining Maintenance Logs with Case-Based Reasoning.*

Mark Devaney, Ashwin Ram, Hai Qiu, Jay Lee, (2005); *Center for Intelligent Maintenance Systems University of Wisconsin-Milwaukee*

<ftp://ftp.cc.gatech.edu/pub/ai/ram/er-05-04.pdf>

- *A data mining based algorithm to enhance maintenance management a medical equipment case study.*

Taha Mokfi, Mahsa Almaeenejad, Mohamad Mohsen Sedighi, *Department of Industrial Engineering Islamic Azad University-South Tehran Branch Tehran, Iran*

http://bpj.academia.edu/MohamadMohsenSedighi/Papers/1491927/A_Data_Mining_Based_Algorithm_to_Enhance_Maintenance_Management_A_Medical_Equipment_Case_Study

Nota: Esta publicación contiene una reseña de otros estudios de *DM* aplicados a mantenimiento y los algoritmos utilizados como referencias.

- *A transformer condition assessment framework based on data mining.*

Yongli Zhu (2005), http://ieeexplore.ieee.org/xpls/abs_all.jsp?amumber=1489207

- *Using Data Mining Techniques in Maintenance of Equipments in Arak Petrochemical Complex.*

Mohammad Ehsanifar *et al*/2011

http://www.eurojournals.com/AJSR_25_12.pdf