



Jose Ignacio González Gómez.
Departamento de Economía Financiera y
Contabilidad
Universidad de La Laguna
www.ecofin.ull.es/users/jggomez.

Tema:

Generalidades de la Minería de Datos

Revisado: 2007

Información extraída de:

<http://www.daedalus.es/AreasMD-E.php>

DAEDALUS - Data, Decisions and Language, S. A

Indice

Indice

1.	Aspectos Generales de la Minería de Datos.....	2
2.	Fases de un proyecto de Minería de Datos.....	2
2.1.	Introducción.....	2
2.2.	Fase Filtrado de Datos.....	2
2.3.	Fase Selección de Variables.....	3
2.4.	Fase Algoritmos de Extracción de Conocimiento.....	3
2.5.	Fase Interpretación y Evaluación.....	4
3.	Casos de estudio.....	4
4.	Bibliografía.....	4

1. Aspectos Generales de la Minería de Datos.

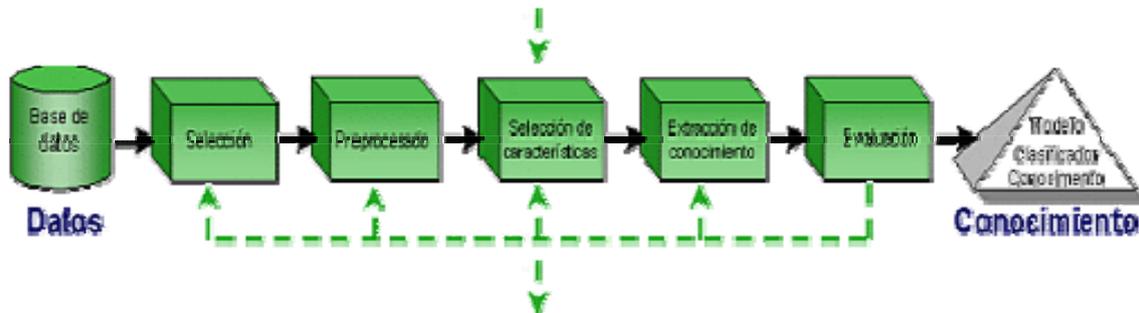
Las técnicas de minería de datos se emplean para mejorar el rendimiento de procesos de negocio o industriales en los que se manejan grandes volúmenes de información estructurada y almacenada en bases de datos. Por ejemplo, se usan con éxito en aplicaciones de control de procesos productivos, como herramienta de ayuda a la planificación y a la decisión en marketing, finanzas, etc.

Asimismo, la minería de datos es fundamental en la investigación científica y técnica, como herramienta de análisis y descubrimiento de conocimiento a partir de datos de observación o de resultados de experimentos.

2. Fases de un proyecto de Minería de Datos.

2.1. Introducción

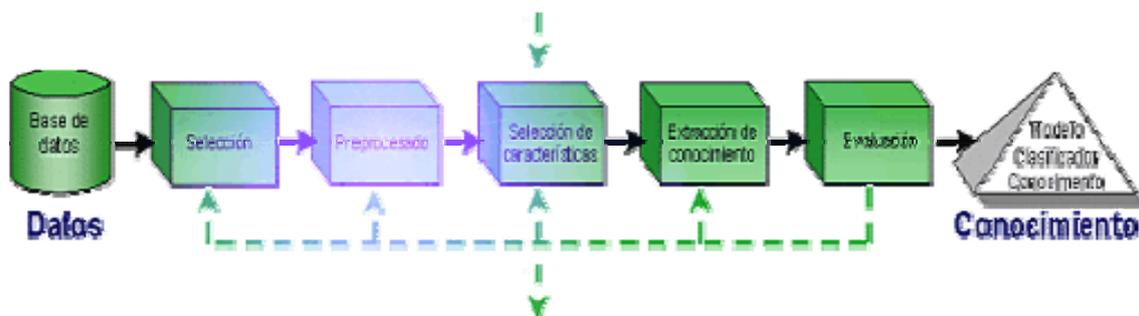
Los pasos a seguir para la realización de un proyecto de minería de datos son siempre los mismos, independientemente de la técnica específica de extracción de conocimiento usada.



El proceso de minería de datos pasa por las siguientes fases:

- Filtrado de datos
- Selección de Variables
- Extracción de Conocimiento
- Interpretación y Evaluación

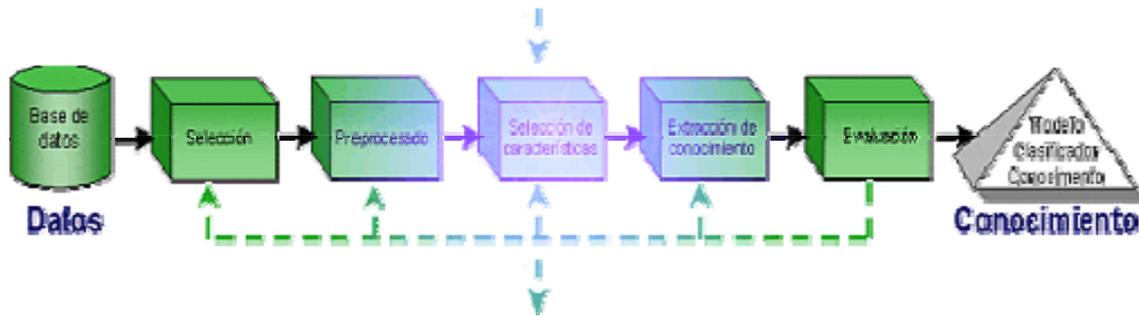
2.2. Fase Filtrado de Datos.



El formato de los datos contenidos en la fuente de datos (base de datos, Data Warehouse...) nunca es el idóneo, y la mayoría de las veces no es posible ni siquiera utilizar ningún algoritmo de minería sobre los datos "en bruto". Mediante el preprocesado, se filtran los datos (de forma que se eliminan valores incorrectos, no válidos, desconocidos... según las necesidades y el algoritmo a usar), se

obtienen muestras de los mismos (en busca de una mayor velocidad de respuesta del proceso), o se reducen el número de valores posibles (mediante redondeo, clustering,...).

2.3. Fase Selección de Variables.

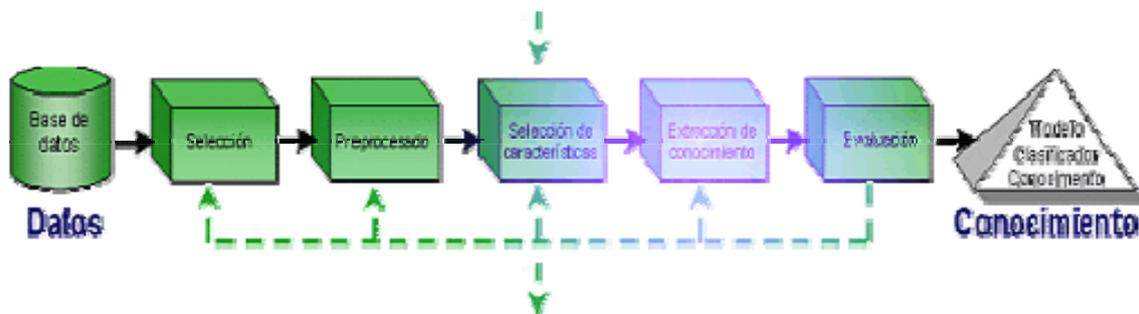


Aún después de haber sido preprocesados, en la mayoría de los casos se tiene una cantidad ingente de datos. La selección de características reduce el tamaño de los datos eligiendo las variables más influyentes en el problema, sin apenas sacrificar la calidad del modelo de conocimiento obtenido del proceso de minería.

Los métodos para la selección de características son básicamente dos:

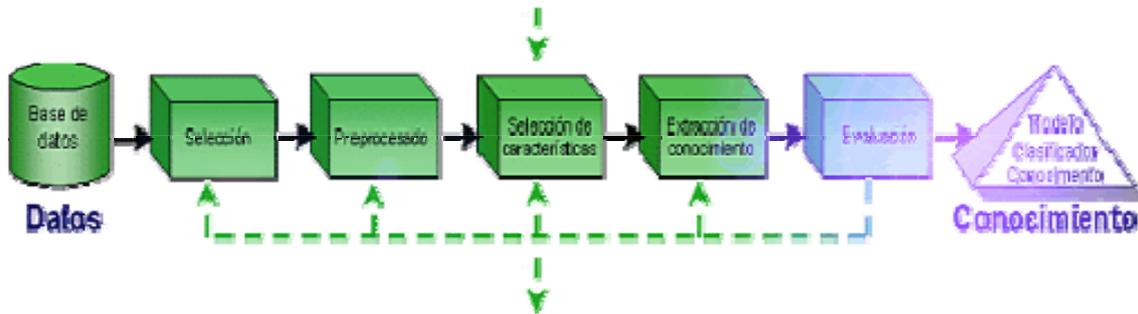
- Aquellos basados en la elección de los mejores atributos del problema,
- Y aquellos que buscan variables independientes mediante tests de sensibilidad, algoritmos de distancia o heurísticos.

2.4. Fase Algoritmos de Extracción de Conocimiento



Mediante una técnica de minería de datos, se obtiene un modelo de conocimiento, que representa patrones de comportamiento observados en los valores de las variables del problema o relaciones de asociación entre dichas variables. También pueden usarse varias técnicas a la vez para generar distintos modelos, aunque generalmente cada técnica obliga a un preprocesado diferente de los datos.

2.5. Fase Interpretación y Evaluación



Una vez obtenido el modelo, se debe proceder a su validación, comprobando que las conclusiones que arroja son válidas y suficientemente satisfactorias. En el caso de haber obtenido varios modelos mediante el uso de distintas técnicas, se deben comparar los modelos en busca de aquel que se ajuste mejor al problema. Si ninguno de los modelos alcanza los resultados esperados, debe alterarse alguno de los pasos anteriores para generar nuevos modelos.

3. Casos de estudio

En la Web de la empresa DAEDALUS (<http://www.daedalus.es/AreasMD-E.php>), se presentan un conjunto de casos entre los que destacan:

<i>Optimización de la producción de cartón en la Industria Papelera</i>	<i>Mejorar el rendimiento de la fabricación de cartón optimizando el control de la velocidad de la línea.</i>
<i>Optimización del Proceso de Producción de Cemento</i>	<i>Desarrollo de un sistema de ayuda a la decisión para la optimización del proceso de producción en una fábrica de cemento.</i>
<i>Aplicación de minería de datos para predicción de ventas</i>	<i>Desarrollo de un modelo para predecir las ventas de un producto en un determinado mes, basándose en datos sobre las ventas en meses previos.</i>
<i>Aplicación de minería de datos para análisis del estado del tráfico en carretera</i>	<i>Desarrollo de un sistema para clasificación del estado del tráfico</i>
<i>Aplicación de minería de datos para predicción de demandas de tiempo de trabajo</i>	<i>Desarrollo de un sistema de predicción del tiempo de trabajo demandado por actividades de reparto postal</i>
<i>Aplicación de minería de datos para segmentación de clientes de entidades financieras</i>	<i>Segmentación de los clientes de un banco mediante un modelo basado en agrupamiento (clustering) borroso.</i>

4. Bibliografía

DAEDALUS - Data, Decisions and Language, S. A (2007): <http://www.daedalus.es/AreasMD-E.php>