



Published on *Dataprix* (<http://www.dataprix.com>)

[Principal](#) > Artículos de Minería de Datos de Dataprix

By *Dataprix*

Created 26/12/2009 - 17:13

Artículos de Minería de Datos de Dataprix

Introducción a la minería de datos

La minería de datos, o data mining, es el proceso no trivial de descubrir patrones válidos, nuevos, potencialmente útiles y comprensibles dentro de un conjunto de datos, según la definición de Piatetsky-Shapiro publicada en la revista "AI Magazine".

Para simplificarlo, podríamos decir que la minería de datos trata de extraer conocimiento a partir de los datos.

Mediante una serie de procesos aplicados en diferentes fases sobre los datos en bruto, y definidos por un experto que conozca el significado de estos datos, y tenga claros los objetivos que persigue, se pueden extraer relaciones entre estos datos, descubrir patrones ocultos y construir modelos que describan este conocimiento.

Las **fases** por las que debería pasar este proceso de descubrimiento de conocimiento son las siguientes:

- **Definición** de la tarea de minería de datos. ¿Qué objetivos se persiguen?
- **Selección** de los datos
- **Preparación** de los datos
- Aplicación de **procesos** de minería de datos sobre los datos preparados
- **Evaluación** y interpretación del modelo obtenido
- **Integración** de los resultados en los sistemas de información

Es un proceso continuo, y que puede constar de diferentes iteraciones, donde los resultados de una iteración alimentan el inicio de la siguiente.

Por supuesto, para la realización de todo el proceso existen diferentes herramientas especializadas que facilitan, o posibilitan, el paso por todas las fases. Dos de las más conocidas son *SAS Enterprise Miner* y *SPSS Clementine*. También existe algún proyecto de software de libre distribución, como *WEKA*, desarrollado en la Universidad

de Waikato, que permite realizar procesos de Data Mining.

El modelo CRISP-DM de minería de datos

En el post de nuestro foro [CRISP-DM Traducción al español](#) ^[1] Daniel Alejandro adjunta un documento con una traducción en castellano de la metodología de CRISP-DM para el desarrollo de modelos de minería de datos.

El contenido del mismo nos parece tan interesante para la comunidad que lo publicamos en forma de manual online para hacerlo más accesible a todos los miembros. Se puede consultar siguiendo el enlace [Metodología CRISP-DM 1.0](#) ^[2]

Adjuntamos también el documento original en inglés a partir del cual Daniel ha realizado la traducción.

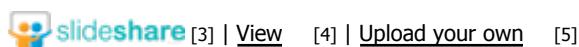
Caso de estudio: Business Intelligence aplicado a la banca

En estas diapositivas se presenta el caso de estudio de un banco que se plantea la necesidad de un mayor **conocimiento de sus clientes** para poder definir adecuadamente sus **estrategias de negocio**.

Gracias a la utilización de herramientas de **Business Intelligence**, en concreto de **Data Warehouse** y **Data Mining**, y a la definición de unos claros *objetivos de negocio*, este banco pudo analizar el comportamiento de sus clientes, segmentarlos, tomar decisiones estratégicas en función de este comportamiento, realizar predicciones y analizar los resultados de la aplicación de estas decisiones, valorando así el retorno de la inversión.

La presentación está estructurada en los siguientes apartados:

- **La economía del negocio y la gestión de clientes**
- **Modelamiento de valor - Modelamiento de potencial**
- **Segmentación - Modelamiento de deserción**
- **Aplicaciones prácticas**
- **Resultados**



La minería de datos vista como una tortura

En el artículo [Data Mining: Torturando a los datos hasta que confiesen](#) [6] Luis Carlos Molina proporciona una visión muy clarificadora sobre la minería de datos, incluyendo interesantes ejemplos de aplicaciones de la misma.

Incluyo el resumen del artículo y el índice, extraídos de la misma publicación:

Resumen: El título de este artículo es una explicación informal de la actividad que realiza una tecnología denominada data mining (minería de datos). Lo que se pretende con esta tecnología es descubrir conocimiento oculto a partir de grandes volúmenes de datos. Desde la década pasada, debido a los grandes avances computacionales, se ha ido incorporando a las organizaciones para constituirse en un apoyo esencial al momento de tomar decisiones. Organizaciones tales como empresas, clubes profesionales deportivos, universidades y gobiernos, entre otros, hacen uso de esta tecnología como ayuda en la toma de sus decisiones. Algunos de estos ejemplos serán citados en el presente trabajo.

SUMARIO

1. Introducción
2. Data mining: conceptos e historia
3. Aplicaciones de uso
- 3.1. En el gobierno

- 3.2. En la empresa
- 3.3. En la universidad
- 3.4. En investigaciones espaciales
- 3.5. En los clubes deportivos
- 4. Extensiones del data mining
- 4.1. Web mining
- 4.2. Text mining
- 5. Conclusiones

Para tener una visión más gráfica sobre este contenido, se puede consultar la presentación [Data mining: Torturando los datos hasta que confiesen](#) [7], realizada por el mismo autor.

Herramientas de Data Mining

En la publicación electrónica MCData.ti se puede encontrar una clasificación bastante completa de diferentes herramientas relacionadas con el business intelligence y la gestión de datos.

Esta es la descripción que se realiza de las herramientas de Minería de datos

Empresa: Computer Associates

[Producto: CleverPath Predictive Analysis Server](#)

Descripción: Solución que puede analizar y extraer conocimiento crítico a partir de los datos, descubriendo los factores significativos, causas y cambios que influencian en el éxito del negocio. Posteriormente, los utiliza para aplicar inteligencia en tiempo real en las aplicaciones e-business para adaptar y ofrecer servicios personalizados a clients, proveedores, partners y empleados.

Empresa: MicroStrategy

[Producto: MicroStrategy Desktop](#)

Descripción: Software corporativo que ofrece funciones de consulta y generación de informes, análisis y workflow de soporte para la toma de decisiones. También incluye amplias capacidades de análisis online de los datos corporativos.

Empresa: MIS

[Producto: MIS Delta Miner](#)

Descripción: Herramienta concebida para efectuar análisis de alto nivel, detección de desviaciones y análisis interactivo sobre múltiples fuentes, tanto transaccionales (SQL Server, Oracle, DB2, etc) como multidimensionales (Hyperion Essbase, SAP BW, MS Analysis Services y Oracle Express).

Empresa: Oracle

[Producto: Oracle 9i Data Mining](#)

Descripción: Función de Oracle 9i Enterprise Edition que permite diseñar aplicaciones de BI que más tarde realizan funciones de «minería» en las bases de datos corporativas para descubrir nueva información e integrarla con las aplicaciones de negocio.

Empresa: SAS

[Producto: SAS Enterprise Miner](#)

Descripción: Solución de minería de datos que permite incorporar patrones inteligentes

a los procesos de marketing, tanto operativos como estratégicos.

Producto: SAS Text Miner

Descripción: Solución que permite la extracción de patrones de conocimiento comunes desde una gran variedad de documentos de texto, incluyendo informes de llamadas comerciales, expedientes médicos y comunicaciones de clientes vía correo electrónico.

Empresa: SPSS

Producto: Clementine

Descripción: Sistema integrado de minería de datos que permite encontrar patrones en la información para facilitar la toma de decisiones a los usuarios. Funciona sobre todas las plataformas hardware y sistemas operativos, incluyendo Unix, VMS y Windows NT.

Empresa: Teradata

Producto: Teradata Warehouse Miner

Descripción: Solución que facilita la construcción de modelos analíticos directamente sobre la base de datos, eliminando así la necesidad de extraer muestras a sistemas del exterior. Además permite analizar datos sin cambiarlos de sitio y es capaz de visualizar la información para representar regresiones lineales o logísticas.

Source URL: <http://www.dataprix.com/articulos-miner-datos-dataprix>

Links:

- [1] <http://www.dataprix.com/node/104>
- [2] <http://www.dataprix.com/CRISP-DM>
- [3] <http://www.slideshare.net/?src=embed>
- [4] <http://www.slideshare.net/customersforever/caso-de-estudio-banca>
- [5] <http://www.slideshare.net/upload>
- [6] <http://www.uoc.edu/web/esp/art/uoc/molina1102/molina1102.html>
- [7] <http://www.lsi.upc.es/~lcmolina/SC/html/paper/chih-dm.pdf>