

Instituto Tecnológico Nacional de México

Maestría en Ingeniería Administrativa



MINERIAS DE DATOS

Fundamentos de Ingeniería Administrativa



Por: Sánchez Ávila María Fernanda

Abril 2016

Índice

Introducción.....	1
Generalidades.....	2
¿Qué es la minería de datos?	3
Aplicación de las minerías de datos.....	3
Principales características y objetivos de la Minería de Datos.....	4
Etapas del Proceso del Datamining	6
Determinación de los objetivos	6
Reprocesamiento de los datos	6
Determinación del modelo	6
Análisis de los resultados	7
Aplicaciones de Uso.....	7
En el Gobierno.....	7
En la empresa	8
En la Universidad.....	8
En Investigaciones Especiales	8
En clubes deportivos	8
Conclusión	9
Propuesta de tema de tesis.....	10
Objetivo General	10
Agradecimientos	10
Referencias.....	11
Ilustración 1 Fases del proyecto del datamining (Vallejos, 2006).....	7

Introducción

Algo que suele ser peor que no tener información disponible es tener muchísima y no saber cómo manejarla. Durante los últimos años, la humanidad ha desarrollado una gran habilidad y capacidad para generar y recolectar datos, debido a que existen las máquinas que pueden procesarla a un costo bajo de almacenamiento.

Pero sin embargo, existen dentro de esas cantidades enormes de datos mucha información oculta y que es de gran importancia pero a la que no se puede acceder mediante las técnicas clásicas de recuperación de información.

Todo el descubrimiento de esa información es posible gracias a las minerías de datos o DataMining, que entre otras de sus características que veremos en el desarrollo de este artículo se encuentra el uso de la inteligencia artificial que permite identificar los patrones y relaciones en los datos permitiendo la creación de modelos.

Durante el desarrollo de este artículo se desarrollará el tema de las minerías de datos, cuáles son sus características, metodología y sus principales áreas de aplicación.

Palabras clave:

- Datamining
- Minerías de datos
- Análisis de información
- Procesamiento de datos
- Bases de datos

MINERIAS DE DATOS

Generalidades

Resulta curioso que en la era que actualmente vivimos, considerada la era de la información ya que gracias a ella podemos enterarnos rápidamente de los acontecimientos alrededor del mundo, los datos pueden llegar a formar una pared confusa y abigarrada¹.

La materia prima que sirve para tomar las decisiones no es siempre la más asequible, resulta entonces necesario salir a buscarla y aunque parezca sencillo no lo es, ya que es preciso extraer datos de manera selectiva para que se logre obtener un beneficio económico.

A todo esto se le conoce como minería de datos, aunque se definirá más adelante de manera correcta, es como decir que si los enanos de Blanca Nieves salían a trabajar para encontrar piedras preciosas, también tenían que tomar los picos y las palas idóneas para conseguirlas, en este caso sería para conseguir la información correcta.

El objetivo principal de la minería de datos se trata de extraer la información de un conjunto de datos, de trabajarla y pulirla para poder transformarla en una estructura que sea comprensible para usarla posteriormente.

Las organizaciones que emplean la minería de datos pueden ver rápidamente el retorno de su inversión puesto que dejan de dar pasos equivocados, una aplicación es la detección de los hábitos de consumo en un supermercado.

(Durán Mena, 2014)

¹ Abigarrada: Heterogéneo, reunido sin incierto. (RAE, 2014)

¿Qué es la minería de datos?

A continuación se presentan algunas definiciones de minerías de datos de varios autores:

- La minería de datos es el proceso que permite detectar la información adicional de los grandes conjuntos de datos debido a que utiliza un análisis matemático que le permita deducir los patrones y las tendencias que existen en los datos. (SQL Server, 2014)
- La minería de datos es aquel proceso que tiene como propósito extraer, descubrir y almacenar cierta información que sea relevante de amplias bases de datos, por medio de programas de búsqueda y otros indicadores que tienen una explicación y que puedan descubrirse mediante la aplicación de estas herramientas. (Larrieta & Santillán Gómez, 2007)
- La minería de datos también conocida como «descubrimiento de datos o de conocimiento» y es el proceso para analizar los datos desde distintos puntos de vista y resumiéndolos en información útil. (Tecnologías de Información, 2009)
- El datamining o minería de datos, es un conjunto de técnicas y de tecnologías que permiten explorar las grandes bases de datos, manera automática o también semiautomática, con el propósito de encontrar patrones que sean repetitivos, tendencias o reglas que permitan explicar el comportamiento de datos en un contexto determinado. (Sinnexus, 2007)

Aplicación de las minerías de datos

(SQL Server, 2014) Los modelos de las minerías de datos, se pueden aplicar en los siguientes escenarios:

- **Previsión:** Permite calcular las ventas y predecir cargas o tiempos de inactividad de los servidores.
- **Riesgo y probabilidad:** Ayuda a elegir a los mejores clientes para la correcta distribución de correo y asigna probabilidades de diagnóstico o algunos otros resultados.
- **Recomendaciones:** Sirve para determinar productos que se pueden vender juntos y generar algunas recomendaciones.
- **Buscar secuencias:** Analiza artículos que clientes han introducido en un carrito de compra y así predecir posibles eventos.
- **Agrupación:** Separa clientes o eventos en clústeres determinados y así analizar o predecir afinidades.

Principales características y objetivos de la Minería de Datos

(Vallejos, 2006) Las características más importantes de las minerías de datos son:

- Explorar datos que se localizan en las profundidades de las bases de datos, o almacenes de datos ya que suelen almacenar mucha información a través del tiempo.
- En ciertos casos, esas bases o almacenes de datos se convierten en mercados de datos o se suelen mantener en servidores del Internet o del Intranet.
- El ambiente externo de la minería suele ser la relación servidor – cliente.
- Las herramientas del datamining ayudan a extraer mineral de información que está enterrado en archivos corporativos o en los registros públicos.

- La minería de datos produce algunos tipo de información:
 - Asociaciones
 - Clasificaciones
 - Secuencias
 - Pronósticos
 - Agrupamientos

En la minería de datos se seleccionan los datos esperando que de ellos emerjan algunas hipótesis y se busca que los datos describan o indiquen por qué son como son.

Posteriormente, se valida la hipótesis y de ahí la minería de datos debe de presentar un enfoque exploratorio, aunque usar el datamining para confirmar hipótesis es algo peligroso ya que se hace una inferencia un poco válida.

El datamining es una tecnología que está conformada por etapas que integran varias áreas pero que no deben de confundir con un software.

Actualmente existen algunas aplicaciones o herramientas de las minerías de datos que son poderosas y que facilitan el desarrollo de proyectos aunque normalmente se complementan con otra herramienta.

Etapas del Proceso del Datamining

Aunque las minerías de datos suelen ser distintas entre sí, el proceso común de ellas se compone de cuatro etapas principales:

Determinación de los objetivos

Esta primera etapa trata de la delimitación de objetivos que el cliente puede desear bajo la orientación de un especialista del datamining.

Reprocesamiento de los datos

En la segunda etapa, se refiere básicamente a la selección, limpieza, enriquecimiento, reducción y transformación de bases de datos, esta etapa generalmente consume alrededor de un setenta por ciento del tiempo total de un proyecto de datamining.

Determinación del modelo

Tercera etapa, en esta se comienza realizando un análisis estadístico de datos y después se visualiza de manera gráfica para tener una aproximación.

Según objetivos planteados y las tareas que se deben de llevar a cabo, pueden utilizarse algunos algoritmos desarrollados en distintas áreas de Inteligencia Artificial.

Análisis de los resultados

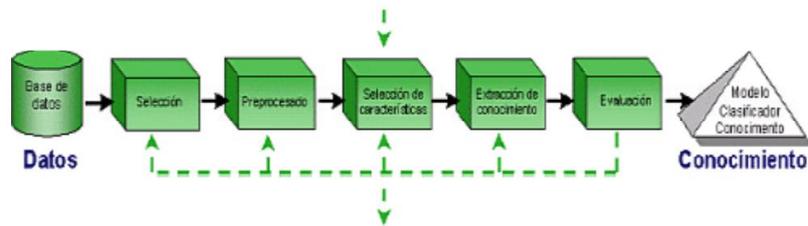


Ilustración 1 Fases del proyecto del datamining (Vallejos, 2006)

Durante la última etapa, se verifican los resultados obtenidos y se cotejan con los análisis estadísticos y las gráficas.

El cliente debe de decidir si son novedosos y si estos aportan un conocimiento nuevo que les permitan tomar decisiones.

Aplicaciones de Uso

Durante cada año, en diferentes congresos y talleres se reúnen investigadores con distintas aplicaciones, sobre todo en Estados Unidos, la minería de datos se ha ido incorporando a la vida de organizaciones, universidades, gobiernos, hospitales y diversas empresas están interesadas en explorar sus bases de datos.

En el Gobierno

El FBI analizará bases de datos comerciales para poder detectar a terroristas.

En la empresa

- Permite la detección de fraudes en tarjetas de crédito.
- Descubren porque las personas desertan de una compañía de telefonía móvil.
- Identifican hábitos de compra en los supermercados
- Predicen el tamaño de audiencias televisivas

En la Universidad

Permite conocer si los recién titulados de una universidad llevan a cabo actividades profesionales que estén relacionadas con lo que estudiaron.

En Investigaciones Especiales

Desarrollo del proyecto SKYCAT, el cual se basa en técnicas de agrupación y árboles de decisión para poder clasificar objetos con alta confiabilidad.

En clubes deportivos

Equipos de la NBA utilizan aplicaciones que son inteligentes para apoyar a su equipo de entrenadores.

(Vallejos, 2006)

Conclusión

El datamining o minerías de datos como vimos en el desarrollo de este artículo, sirven para cultivar la lealtad de los clientes ya que permite ofrecerle algo que perciben como valioso, ya que una de sus características es identificar patrones de conducta con cierta tendencia a darse de baja en base a datos de aquellos clientes que ya lo hicieron, así las organizaciones siempre pueden estar un paso adelante y ofrecer algunos incentivos para retener a sus clientes.

Existen también muchas áreas de aplicación importantes para este tipo de análisis de la información como la medicina, prevención y control de fraudes, investigación de actos que están vinculados con el terrorismo, ingeniería y genética.

Las personas que se dedican a las minerías de datos dicen que se trata básicamente de estadística mezclada con negocios y sostienen que los métodos que utiliza y el tipo de problemas que pueden enfrentar las hace únicas y sumamente relevantes.

En resumen, las minerías de datos se presentan como una tecnología que está emergiendo con varias ventajas por supuesto, como el punto de encuentro entre investigadores y personas de negocios, y el ahorro de grandes cantidades de dinero a la organización además de que permite abrir nuevas oportunidades de negocio. Además de que trabajar con el datamining implica cuidar tantos detalles que al final permite la toma de decisiones de manera precisa.

Propuesta de tema de tesis

Aplicación de minerías de datos para el análisis del comportamiento de ventas de una organización.

Objetivo General

Aplicar técnicas de minerías de datos para analizar el comportamiento o patrones de conducta de clientes de una organización.

Agradecimientos

Al Tecnológico Nacional de México por ser mi alma máter y al Dr. Fernando Aguirre y Hernández por su apoyo y motivación para realizar estos artículos en la materia de Fundamentos de Ingeniería Administrativa.

Referencias

- Durán Mena, C. (6 de Agosto de 2014). *Forbes México*. Obtenido de <http://www.forbes.com.mx/mineria-de-datos-informacion-precisa-y-relevante/>
- Larrieta, M. I., & Santillán Gómez, A. M. (2007). *EJournal UNAM*. Recuperado el Marzo de 2016, de Minería de datos: Concepto, características, estructura y aplicaciones: <http://www.ejournal.unam.mx/rca/190/RCA19007.pdf>
- RAE. (2014). *Real Academia Española*. Obtenido de <http://dle.rae.es/srv/search?m=30&w=abigarrado>
- Sinnexus. (2007). *Business Intelligence Informática estratégica*. Recuperado el Marzo de 2016, de : http://www.sinnexus.com/business_intelligence/datamining.aspx
- SQL Server. (2014). *Microsoft*. Obtenido de [https://msdn.microsoft.com/es-mx/library/ms174949\(v=sql.120\).aspx](https://msdn.microsoft.com/es-mx/library/ms174949(v=sql.120).aspx)
- Tecnologías de Información. (2009). *Sistemas de Información: Procesamiento de datos, planificación y gestión de recursos*. Recuperado el Marzo de 2016, de <http://www.tecnologias-informacion.com/mineria-de-datos.html>
- Vallejos, S. J. (2006). *ExaUnne.edu*. Recuperado el Marzo de 2016, de http://exa.unne.edu.ar/informatica/SO/Mineria_Datos_Vallejos.pdf