

Minería de Datos como soporte a la toma de decisiones empresariales

*Yelitza Josefina Marcano Aular
y Rosalba Talavera Pereira*

*Universidad del Zulia, Núcleo Punto Fijo.
ymarcano@hotmai.com; talavera_p@hotmail.com*

Resumen

La tarea por mejorar el acceso a la información está cobrando cada vez más fuerza, especialmente en los negocios actuales, donde se requiere principalmente de procesos basados en el recurso información, de manera automatizada y reutilizable. En ese orden de ideas, este artículo constituye una primera aproximación al área de la Minería de Datos y tiene como objetivo examinar y describir las técnicas y herramientas que emergen en esa área de investigación, apoyándose para ello en una reflexión teórica-cualitativa que contribuya a un mayor entendimiento del alcance y limitaciones de la Minería de Datos como soporte a la toma de decisiones empresariales. Entre los beneficios que ofrece la técnica están la posibilidad de elevar los niveles de competencia de los negocios, basándose en la rapidez para identificar, procesar y extraer la información que realmente es importante, descubriendo conocimiento y patrones en bases de datos. Su facilidad de uso hace que se pueda aplicar a cualquier área del conocimiento. Como limitaciones destacan la necesidad de dedicar mucho esfuerzo al establecimiento de medidas de evaluación del resultado derivado de la aplicación de la minería, así como el desafío que representa analizar datos que cambian en tiempo real.

Palabras clave: Minería de datos, toma de decisiones empresariales, descubrimiento de conocimiento, bases de datos.

Data Mining as a Support for Enterprise Decision Making

Abstract

The task of improving access to information is receiving more and more attention, specially in modern business, where processes based on information as a resource in an automatic and re-usable manner, are required. In this sense, this article constitutes a first approximation to the area of Data Mining and its objective is to examine and describe the techniques and tools emerging in this area of research, based on a theoretical-qualitative reflection that contributes to a greater understanding of the potential and limitations of Data Mining as a support for enterprise decision making. Among the benefits that the technique offers is the possibility of elevating competence levels in businesses, based on a quick identification of, processing and extraction of information that is truly important, as well as discovering knowledge and patterns in data bases. Its ease of use makes it applicable to any area of the knowledge. As limitations, the need to dedicate effort to the establishment of measures of evaluation of the result derived from the application of the data-mining process, as well as the challenge that represents the analysis of data that change in real time are among the most outstanding.

Key words: Data Mining, enterprise decision making, discovery of knowledge, data bases.

INTRODUCCIÓN

En la actualidad las organizaciones suelen moverse dentro de estructuras identificadas con un cambio continuo; por ello, las empresas privadas tanto como las públicas deben tener la capacidad de ser adaptativas, aprender cómo resolver problemas y generar conocimiento, para establecer nuevos métodos en pro de la resolución de los mismos.

Las organizaciones, en la búsqueda por la obtención de los mejores resultados de su gestión organizacional, adoptan la flexibilización como estrategia, con el objetivo de adecuarse a un mercado globalizado, dando origen a un proceso que incide en su sistema estructural. Así pues, una empresa flexible es la que se orienta hacia los clientes, posee tecnología

nueva y presenta acuerdos laterales de organización e innovación (Hansen y Mouritsen, 1999).

Las aplicaciones necesarias para gestionar el flujo de información en las actividades de negocio se pueden clasificar en dos importantes categorías: las aplicaciones que manejan las transacciones y las estadísticas que ayudan a convertir los datos en información útil para la toma de decisiones. Además está el sistema de indicadores, formado por las bases de datos donde se almacenan los datos importantes para evaluar y mejorar el funcionamiento de las actividades que componen la cadena de suministro y por aplicaciones de análisis que facilitan la comprensión de las tendencias y patrones presentes en los datos. El sistema de indicadores se considera como un instrumento de integración básico a través de la comunicación y diálogo que se establece, en base a los datos, entre los diferentes actores del proceso.

En la visión de Castañeda y Rodríguez (2003), el uso de la Minería de Datos o Data Mining, como soporte a las decisiones en las actividades de negocio, requiere mucho más que la aplicación de sofisticadas técnicas como redes neuronales o árboles de decisión sobre las tablas de datos. Por esta razón, en el presente documento se muestra a la Minería de Datos por un lado, como uno de los pasos del proceso de descubrimiento de conocimiento en base de datos (KDD) y por otro lado como un proceso que consta de diferentes fases, en las cuales se utilizan como apoyo, técnicas relacionadas con la estadística, el reconocimiento de patrones y algoritmos de aprendizaje, entre otras.

Todos estos estudios han incrementado el deseo desenfrenado por demandar un mayor control de los procesos u operaciones y servicios, visto como núcleo de una gestión global, fundamental para proporcionar servicios de calidad y lograr un rendimiento óptimo de las inversiones, en infraestructuras comerciales, en un entorno competitivo dirigido hacia una gestión de clientes.

Este trabajo constituye un primer acercamiento a un área de investigación de reciente data, el cual tiene como propósito presentar algunas bases teóricas sobre la incidencia de la Minería de Datos como soporte en la toma de decisiones, aplicadas a las actividades de negocio. La elaboración de la reflexión teórica hace énfasis en los postulados metodológicos del paradigma cualitativo, el cual permite la construcción del conocimiento partiendo de una visión integral, interpretativa y contextual

del fenómeno a estudiar. Las teorías consultadas se interpretaron para establecer por inferencia deductiva algunas consideraciones relacionadas a la Minería de Datos y a algunos indicadores que permitan medir el interés y el impacto del conocimiento que se puede obtener, al emplearla, como soporte para la toma de decisiones en las organizaciones.

1. EL TRATAMIENTO DE LOS DATOS PARA LA TOMA DE DECISIONES EN LAS ACTIVIDADES DE NEGOCIO

Desde un punto de vista más pragmático y asociándolo directamente a las Actividades de Negocios, la Minería de Datos es el conjunto de metodologías, aplicaciones y tecnologías que permiten reunir, depurar y transformar datos de los sistemas transaccionales e información no estructurada (interna y externa a la compañía) en información estructurada, para su explotación directa o para su análisis y conversión en conocimiento y así dar soporte a la toma de decisiones sobre el negocio. Ahora bien, Piatesky-Shapiro (1991) destacan que desde un punto de vista más teórico, la Minería de Datos se define como el proceso completo de extracción de información, que se encarga además de la preparación de los datos y de la interpretación de los resultados obtenidos, a través de grandes cantidades de datos, posibilitando de esta manera el encuentro de relaciones o patrones entre los datos procesados.

Por su parte, Molina y García (2004) explican que los datos tal cual se almacenan en las bases de datos no suelen proporcionar beneficios directos; su valor real reside en la información que podamos extraer de ellos, es decir, información que nos ayude a tomar decisiones o a mejorar la comprensión de los fenómenos que nos rodean. Ejemplos de ello pueden ser: contrastar que todo va bien, analizar diferentes aspectos de la evolución de la empresa, presentar información de forma más intuitiva, comparar información en diferentes períodos de tiempo, comparar resultados con previsiones, identificar comportamientos y evoluciones excepcionales, confirmar o descubrir tendencias e interrelaciones, entre otras acciones.

Para tal efecto, se deben definir medidas cuantitativas para los patrones obtenidos (precisión, utilidad y beneficio obtenido), para establecer medidas de interés que consideren la validez y simplicidad de los patrones obtenidos mediante alguna de las técnicas de Minería de Datos. El

objetivo final de todo esto es incorporar el conocimiento obtenido en algún sistema real, tomar decisiones a partir de los resultados alcanzados, o simplemente registrar la información conseguida y suministrársela a quien esté interesado.

Cabe resaltar que los efectos sobre la eficiencia de los resultados en la Minería de Datos como apoyo a la toma de decisiones está generando miles de opiniones desde diferentes perspectivas, entre ellas se puede destacar: la imposibilidad de encontrar conclusiones únicas referentes a la evaluación de un resultado obtenido con esta técnica. Sin embargo, la consecuencia de los datos arrojados es interesante, por algo más que su precisión.

En el siguiente artículo se realizó un análisis de la aplicación del Data Mining en los negocios, haciendo énfasis en cuatro indicadores propuestos por los autores, Castañeda y Rodríguez (2003), los cuales esbozan el atractivo que ofrece la Minería de Datos en la implementación de las actividades de negocio: tales como la bondad, aplicabilidad, la relevancia y la novedad; indicadores que aportan una idea de las implicaciones y utilidades que proporciona esta práctica.

a) Indicadores de la bondad del resultado

Los índices de bondad de resultado tratan de aportar una idea acerca del error que se comete al emplear un modelo para realizar una tarea. Tal como manifiestan Padmanabhan y Tuzhilin (1999), ésta es una medida de la fortaleza estadística del resultado. Para este indicador se utilizan las siguientes medidas: Precisión, Ratio de error, Varianza y Matriz de confusión, siendo las dos últimas derivaciones de las anteriores. La precisión se utiliza cuando el resultado se presenta en forma de clasificación o estimación, la cual se mide a través del porcentaje de predicciones que son correctas. Para efectos de la clasificación, se emplea el porcentaje de casos bien clasificados y para la estimación del porcentaje de registros, se emplea una estimación que el decisor considere correcta. Para medir la precisión se puede emplear el coeficiente de confianza, el cual no es más que la probabilidad condicionada de un hecho con respecto a otro.

La distancia es otra técnica de Minería de Datos empleada cuando se disponen de variables continuas y numéricas, mediante la raíz cuadrada de la suma al cuadrado de las distancias en cada eje. Una medida que

complementa a la precisión es el Ratio de error, que mide el porcentaje de casos en los que el resultado no coincide con la realidad.

b) Indicadores de relevancia del resultado

Los indicadores más representativos en este grupo son el Coeficiente de cobertura, el Coeficiente de apoyo y el Coeficiente de significación. Estos indicadores tienen que ver directamente con la importancia que tiene el resultado arrojado por las técnicas de minería y miden la aportación a la situación actual y la frecuencia de utilidad del resultado, cuando la presentación de éstos se hace en forma de reglas.

El Coeficiente de cobertura mide el porcentaje de registros en los cuales se puede aplicar la regla. Por otro lado, el Coeficiente de apoyo permite mostrar el porcentaje de ocasiones en que globalmente aparece la relación descrita por la regla, se recomienda representar el resultado en porcentaje. Por último, el Coeficiente de significación sirve para medir el grado de importancia de la regla a través de la aportación que supone respecto a la pura probabilidad.

c) Indicadores de novedad del resultado

Cuando la información es excesivamente abundante y obvia, puede presentarse el problema al generar reglas. Para ello, existe el Coeficiente de novedad, creado para indicar si una regla es interesante o no en función del número de reglas ya generadas, para un área de conocimiento concreta. Su objetivo es ayudar a evitar las redundancias en su obtención. Autores como Buchner et al. (1999), entre otros, abogan por la inclusión del conocimiento previo del negocio, e intuición que detentan las decisiones para de esta manera: restringir el espacio de búsqueda, obtener conocimiento más preciso y eliminar aquél que resulte no interesante.

d) Indicadores de aplicabilidad del resultado

La dinámica de las organizaciones actuales demanda cada vez más, tiempos de respuesta más rápidos, por lo cual es necesario que tanto la creación o generación de modelos como los resultados del mismo, deben estar disponibles en el menor tiempo posible. Para lograr esto, hay que buscar la simplicidad de los modelos y de la forma de representar la salida o resultados del análisis, para transformar el conocimiento obtenido y poder aplicarlo al negocio; para lograr esto, se cuenta con el Coeficiente de Simplicidad, la Tasa Interna de Retorno y el Valor Actual Neto.

2. LA RENTABILIDAD ECONÓMICA DE LAS EMPRESAS A TRAVÉS DE LOS PROCESOS MINERÍA DE DATOS, COMO HERRAMIENTA INTELIGENTE COMPETITIVA

Es innegable que organizaciones comerciales, analistas de inversiones, entidades de crédito y demás usuarios especializados, cada vez solicitan más información sobre la rentabilidad de todas las actividades de negocio de la organización. ¿Están capacitados los solicitantes para procesar correctamente todo ese recurso de datos, cada vez más voluminoso y complejo? Nos inclinamos a pensar que no, por la siguiente razón: El volumen creciente de datos a que se enfrentan los usuarios, especialmente aquellos especializados en economía, finanzas y marketing, está haciendo que la sobrecarga informativa convierta los datos en una masa en la cual no es posible ver con claridad los ámbitos y complejos procesos de la situación y perspectivas de la empresa. Ante este panorama nos preguntamos ¿Qué medidas pueden adoptarse? Cabe indicar que los avances tecnológicos han conducido a un mundo complejo y cambiante, pero contradictoriamente es una fuente de soluciones.

Siguiendo en el mismo orden de ideas, se resalta que la Minería de Datos o explotación de datos ha revolucionado y potenciado las bases de datos tradicionales, pues permite identificar mediante la aplicación automática de algoritmos recursivos, las variables más relevantes para una decisión concreta o los subconjuntos de entidades (empresas, clientes, entre otros) más homogéneos con arreglo a unas características determinadas.

Otro importante hecho que describen Molina y García (2004) es que con la implementación de procesos de Minería de Datos, a través de la aplicación de algunas técnicas estadísticas avanzadas y nuevos métodos de extracción de conocimiento en grandes base de datos, se pueden determinar las características contables de las empresas más rentables, al igual que el perfil de sus clientes. Se hace imprescindible, por un lado, un análisis exploratorio profundo de la base de datos y el empleo de métodos robustos, que hagan que dichos componentes sean menos sensibles a los amplios casos estadísticos. Por otro lado, es aconsejable diseñar con base a opiniones de expertos, si no hay información adecuada, o utilizar algún sistema de aprendizaje, por ejemplo, la utilización de redes neuronales, para el descubrimiento de patrones y extraer la información de la base de datos disponible.

En fin, estos métodos y procedimientos se han convertido en retos tecnológicos para procesar los datos y convertirlos en conocimiento útil para la toma de decisiones. Este camino se presenta como una opción para las organizaciones que quieran ser competitivas, valiéndose de la experiencia acumulada, la cual sin duda alguna constituye el principal activo del que se dispone para la creación de valor. De esta manera, una organización que reflexiona, documenta y aprende, está en condiciones de innovar y obtener ventajas competitivas.

3. INFLUENCIA DE LA TECNOLOGÍA PARA LA CONSOLIDACIÓN DE LAS ACTIVIDADES DE NEGOCIOS ESTABLECIDAS EN MINERÍA DE DATOS

Autores como Han y Kamber (2001) consideran que el auge que ha alcanzado actualmente el Data Mining es debido a que en el presente nos encontramos ante enormes cantidades de datos y con la urgente necesidad de transformarlos en información útil y conocimiento. Se dice que sin Data Mining somos “ricos en datos” pero “pobres en información”. El Data Mining es la evolución natural de la tecnología de información. Por lo antes expuesto, se puede inferir que las necesidades de información de las organizaciones ha cambiando a lo largo del tiempo. Hoy día, la creciente dinámica de mercado y competitividad llevan a la necesidad de contar con la información adecuada en el momento indicado y para ello los gerentes necesitan estar bien informados para poder tomar las decisiones de negocio apropiadas. Por otra parte, los datos con los que cuentan dichas organizaciones generalmente se encuentran dispersos a través de diversos sistemas, propiciando de esta manera que datos valiosos se pierdan. De esto último, se puede decir que lo que poseen las empresas son datos, no información.

En términos de lo expuesto, los autores Cativa et al. (2001) manifiestan que los sistemas tradicionales, basados en el procesamiento de transacciones, no brindan una solución adecuada para la satisfacción de los nuevos requerimientos de información de las organizaciones. Esto dio lugar a la aparición de nuevas tecnologías de base de datos conocidas como *Data Warehousing* y OLAP (Procesamiento Analítico On-Line), referido este último a las aplicaciones de bases de datos orientadas a arrays que permiten a los usuarios ver, navegar, manipular y analizar bases de datos multidimensionales, y en general, a lo que se denomina Sis-

temas de Soporte de Decisiones, el cual asiste a los encargados de la dirección de las empresas en la toma de decisiones y proporciona un conjunto de herramientas estructuradas, orientadas a mejorar la efectividad de los resultados de la decisión. Éste se emplea en contextos de decisión no estructurados o semiestructurados.

Siguiendo a Molina y García (2004) se puede decir que dentro de la Minería de Datos se ha tomado la perspectiva orientada a datos, por dos razones. Primero, porque la mayoría de los trabajos en Minería de Datos están enfocados hacia el *Data Warehouse*, referido esto a la arquitectura, algoritmos, herramientas y técnicas utilizadas para agrupar los datos provenientes de múltiples bases de datos, u otras fuentes de información en un repositorio común, sobre el cual se realizarán consultas y análisis. Éste proporciona el apoyo a la Minería de Datos organizando y estructurando los datos; en efecto, tal como lo muestra Morand et al. (2004) se define un nuevo concepto para el almacenamiento de datos, integrando la información generada en todos los ámbitos de una actividad, con una visión orientada al “negocio”, que permite el acceso y explotación de la información contenida en el almacén, facilitando un amplio abanico de posibilidad de análisis multivariados para la toma de decisiones estratégicas.

En este orden de ideas, cabe resaltar la importancia de otras tecnologías de apoyo en los procesos de Minería de Datos, que han sido utilizadas desde hace tiempo, logrando la integración de éstas con la administración de datos, contribuyendo a mejorar la técnica. Thuraisingham (1999) agrega que los métodos estadísticos han propiciado la aparición de varios paquetes de software para computar sumas, promedios y distribuciones, que han ido integrándose con las bases de datos a explorar.

Considerando el criterio de Molina y García (2004), se puede decir que el aprendizaje automático consiste en la obtención de reglas de aprendizaje y modelos de los datos, para lo cual a menudo se necesita la ayuda de la estadística. Por esta razón, los métodos estadísticos y el aprendizaje automático son los dos componentes más importantes de la Minería de Datos. Existen otras tecnologías, entre las que se incluyen la visualización y el procesamiento paralelo, para dar apoyo a la toma de decisiones. Las técnicas de visualización ayudan a presentar los datos, para facilitar la minería de éstos. Por su parte, las técnicas de procesamiento paralelo, a su vez, colaboran para mejorar el rendimiento de la Minería de Datos. Los sistemas de apoyo a la toma de decisiones permiten discriminar los resultados, de manera que puedan ser utilizados en las funciones de dirección en las organizaciones.

Autores como Molina y García (2004) explican cada una de las técnicas que ayudan a la resolución de problemas particulares de la organización, basándose en los datos que éstos poseen. Estas técnicas son: Razonamiento estadístico, Visualización, Procesamiento paralelo, Aprendizaje automático y Apoyo en la toma de decisiones.

- **Razonamiento estadístico**

Las técnicas y métodos estadísticos del razonamiento han sido utilizados durante varias décadas, siendo éstos los únicos medios para analizar los datos en el pasado. Numerosos paquetes estadísticos están ahora disponibles para computar promedios, sumas, y diferentes distribuciones para distintas aplicaciones. Más recientemente, las técnicas estadísticas del razonamiento están jugando un papel importante en la Minería de Datos. Cabe destacar que la Minería de Datos no sustituye la estadística “clásica”, sino que la complementa (Groth, 2000). Así pues, la estadística juega un importante papel en el análisis de los datos, e incluso también en el aprendizaje automático. Debido a esto, no se puede estudiar la Minería de Datos sin un buen conocimiento de la estadística.

- **Visualización**

Las tecnologías de la visualización son buenas para ubicar patrones en un conjunto de datos y pueden ser usadas al comienzo de un proceso de Data Mining, para tomar un *feeling* de la calidad del conjunto de datos. Los modelos de visualización pueden ser bidimensionales, tridimensionales o incluso multidimensionales. Se han desarrollado varias herramientas de visualización para integrarse con las bases de datos ofreciendo una visualización de forma interactiva a la Minería de Datos.

- **Procesamiento paralelo**

El procesamiento paralelo es una técnica que ha sido utilizada durante mucho tiempo. El área se ha desarrollado significativamente, desde sistemas con un único procesador hasta sistemas multiprocesadores. Los sistemas de multiprocesamiento pueden estar formados por sistemas distribuidos o por sistemas centralizados de multiprocesadores con memoria compartida, o con multiprocesadores sin memoria compartida. Estos sistemas no fueron comercializados hasta el desarrollo del Data Warehouse, ya que ellos emplean el procesamiento paralelo para acelerar el proceso de las consultas.

Recientemente estos sistemas se han empezado a utilizar para las aplicaciones comerciales, debido en parte a la explosión del Data Warehouse y de las técnicas de Minería de Datos, donde el rendimiento de los algoritmos de consulta es crítico. Para escalar las técnicas de Minería de Datos se necesita hardware y software apropiado, por lo que los fabricantes de bases de datos están empleando computadores con procesamiento paralelo para llevar a cabo las actividades de minería.

- **Apoyo a la toma de decisiones**

Los sistemas de apoyo a la toma de decisiones son las herramientas que usan los directivos para tomar decisiones eficaces, basándose en la teoría de la decisión. Por su parte, se puede considerar a las herramientas de Minería de Datos como tipos especiales de herramientas de apoyo a la toma de decisiones.

En general, las herramientas de apoyo a la toma de decisiones podrían utilizarse también como herramientas para eliminar los resultados innecesarios e irrelevantes obtenidos de la Minería de Datos. Igualmente pueden ser consideradas de este tipo, herramientas tales como las hojas de cálculo, sistemas expertos, sistemas de hipertexto, sistemas de gestión de información de *Web* y cualquier otro sistema que ayude a analistas y gestores a manejar eficazmente grandes cantidades de datos e información. Recientemente ha aparecido un área nueva llamada gestión del conocimiento, la cual trata de manejar eficazmente los datos, la información y el conocimiento de una organización.

- **Aprendizaje automático**

El aprendizaje automático, en muchos casos, consiste fundamentalmente en el aprendizaje de reglas a partir de los datos y por eso muchas de las técnicas de aprendizaje automático son utilizadas en la actualidad en las actividades de minería. Esta técnica consiste en aprender de las experiencias del pasado con respecto a alguna medida de rendimiento.

Así pues, se han desarrollado distintas técnicas para el aprendizaje automático, incluyendo el aprendizaje conceptual donde se aprende los conceptos desde diferentes ejemplos de entrenamiento, haciendo uso de las redes neuronales, los algoritmos genéticos, los árboles de decisión y la programación de la lógica inductiva. Hay todavía mucha investigación que realizar en esta área, sobre todo en la integración del aprendizaje automático con las diferentes técnicas de gestión de datos. Tal investigación mejorará significativamente el área de Data Mining.

Igualmente Moreno et al. (2002) afirman que la aplicación automatizada de algoritmos de Minería de Datos permite detectar fácilmente patrones en los datos, razón por la cual esta técnica es mucho más eficiente que el análisis dirigido a la verificación, cuando se intenta explorar datos procedentes de repositorios de gran tamaño y complejidad elevada.

En definitiva, Molina y García (2004) dicen que la utilidad de aplicaciones futuras en KDD es de largo alcance. KDD puede usarse como un medio de recuperación de información, de la misma manera que los agentes inteligentes realizan la recuperación de información en la *Web*. Nuevos modelos o tendencias en los datos podrán descubrirse usando estas técnicas. El KDD también puede usarse como una base para las interfaces inteligentes del mañana, agregando un componente del descubrimiento del conocimiento a un sistema de bases de datos o integrando KDD con las hojas de cálculo y visualizaciones.

Resumiendo a Díaz y Pérez (2004), ellos mencionan la importante labor que tienen los algoritmos de Minería de Datos en la exactitud de determinados conjuntos de datos numéricos. Éstos son: los algoritmos de Redes Neuronales e Híbridos de Aprendizaje (inducción de reglas y árboles de decisión), entre otros.

REFLEXIONES FINALES

En este artículo se ha analizado el estado actual de una técnica que está influyendo en las actividades de negocio de las empresas y en la que están implicados un extenso y creciente número de investigadores a nivel mundial, por las implicaciones, estrategias y beneficios que arroja. Como resultado de esta primera aproximación se presentan algunas de las consideraciones más resaltantes derivadas de los planteamientos teóricos, en torno a la Minería de Datos y su impacto en la toma de decisiones en los negocios:

- La Minería de Datos, bien empleada, se convierte en una herramienta estratégica que eleva los niveles de competencia en el cambiante mundo de los negocios. La toma de decisiones efectivas depende de la rapidez con que se identifica y analiza información importante. La existencia de metodologías innovadoras para desarrollar el proceso de identificación y análisis, debe necesariamente mejorar la ventaja competitiva para incrementar el mayor número de clientes.

- Nuestra capacidad para almacenar datos ha crecido exponencialmente los últimos años, pero la capacidad de procesarlos no ha ido a la par. Por tal motivo, es necesario contar con técnicas que tengan la capacidad de procesar y entender datos tanto estructurados como no estructurados, para apoyar la toma de decisiones en cualquier ámbito del conocimiento.
- La Minería de Datos ha tenido una reciente inclusión en los negocios, debido a la enorme preocupación de las empresas por conocer más allá de los datos que éstos manejan.
- Entre las ventajas de la Minería de Datos está su facilidad de uso y la aplicabilidad de un conocimiento adecuado de los distintos tipos de algoritmos empleados, ya que éstos brindan los mismos resultados y cada uno, con una eficiencia diferente. Como desventaja destaca que hay que dedicar mucho más esfuerzo al establecimiento de medidas de evaluación del resultado derivado de la aplicación del Data Mining.
- Para el aprovechamiento de la gran cantidad de conocimiento en la Minería de Datos es necesario reducir la cantidad de datos, quedándonos sólo con la información mínima necesaria, para disminuir el esfuerzo computacional y humano. El resto de la información se vuelve redundante, trayendo consigo ruido y dependencias que deben tratar de evitarse; esto se basa en un axioma fundamental: “la hipótesis más simple”.
- No todos los datos son apropiados para la minería. La búsqueda de patrones debe centrarse en aquéllos que tengan un impacto significativo en el negocio. Si bien los datos de poca utilización se encuentran mezclados con los de alta utilización, contar con un motor de consultas que permita realizar ordenamientos y selección de datos ayuda a determinar cuáles serán aquéllos que se extraerán.
- Un Data Warehouse está diseñado para realizar procesamientos veloces de consultas, lo cual representa una herramienta de suma utilidad en la tarea de identificación del subconjunto de datos requerido.
- Hacer Data Mining sobre datos que se actualizan a menudo es un desafío. Por consiguiente, hay varios problemas que necesitan ser investigados extensamente, antes de que se pueda llevar a cabo lo que se conoce como Data Mining en tiempo real. El uso de indicadores en pro de medir la bondad, aplicabilidad, la relevancia y la

novedad, de los resultados de la Minería de Datos, pueden resultar en ocasiones muy subjetivos, pero los negocios necesitan contar con algún medio que les permita medir el interés y el impacto del conocimiento que se puede obtener al aplicar Minería de Datos. De igual manera, la intervención y experiencia del tomador de decisiones es relevante para establecer algunas medidas y poder calcular los indicadores antes mencionados.

- La tenencia de datos no es el elemento esencial en una toma de decisión acertada. Al convertir dichos datos en información evaluada y ésta en conocimiento para la acción, se proporciona el apoyo necesario para la toma de una decisión argumentada, que oriente a la empresa hacia el cumplimiento de sus metas y objetivos.
- Con Data Mining, las organizaciones cuentan con una nueva forma de ver sus datos, prometiendo beneficios a la solución de una gran variedad de problemas como: planeación económica, inteligencia empresarial, finanzas, análisis de mercados y análisis de perfiles de clientes.

Aunque todavía queda mucho trabajo por hacer en esta temática, se necesita crear más y mejores procesos para generar resultados eficientes en los negocios, y más aún, desarrollar aplicaciones reales que pongan en práctica todos los principios relacionados con la Minería de Datos. En espera de que se alcance esta meta, se han desarrollado y probado ideas a niveles específicos, abriendo nuevos campos para la innovación, suficientemente interesantes para motivar la investigación en esta área.

Bibliografía

- BUCHNER, A.; ANAND, S.; MULVENNA, M.; HUGHES, J. 1999. "Discovery Internet Marketing intelligence through web log mining". Proc. Unicom99. Data Mining & Datawarehousing: Realising the full value of Business Data, pp. 127-138.
- CASTAÑEDA, J. y RODRÍGUEZ, M. 2003. "La Minería de Datos como herramienta de Marketing: Delimitación y Medidas de Evaluación del resultado". Universidad de Granada. Dpto. de Comercialización e investigación de mercados. España.
- CATIVA, S.; D'ABREU, M.; IZQUIERDO, A. y MISLEJ, E. 2001. **Data Warehousing. Ingeniería de Software II.**

- DÍAZ, J. y PÉREZ, R. 2004. "Estado del arte en la utilización de técnicas avanzadas para la búsqueda de información no trivial a partir de datos en los Sistemas de abastecimiento de agua potable". Universidad Politécnica de Valencia. Departamento de Ingeniería Hidráulica y Medio Ambiente. Grupo Mecánica de Fluidos Valencia (España). Disponible en: <http://www.serea.upv.es/serea2004/documentos/modelo.pdf> [Documento on-line], [Consultado 08-06-2005].
- GROTH, R. 2000. **Data Mining building competitive advantage**. Editorial Prentice Hall.
- HAN, J. y KAMBER, M. 2001. **Data Mining, Concepts and Techniques**. Morgan Kaufmann Publishers.
- HANSEN, A.; MOURITSEN, J. 1999. "Managerial technology and Netted Networks: Competitiveness in Action - The work of translating performance in a high tech firm, Organization". Volumen 6, Número 3.
- MOLINA J. y GARCÍA, J. 2004. "Técnicas de análisis de datos". Disponible en: <http://galahad.plg.inf.uc3m.es/~docweb/ad/transparencias/apuntes AnalisisDatos.pdf>. [Documento on-line], [Consultado 18-06-2005].
- MORAND, C.; DAPOZO, G. y OJEDA, C. 2004. **Sistema de Apoyo a las Decisiones para la Gestión Gubernamental de la Producción Agrícola-Forestal en Corrientes**. Dpto. de Informática. Facultad de Ciencias Exactas y Naturales y Agrimensura. Universidad Nacional del Nordeste. Comunicaciones Científicas y Tecnológicas. Número 1449. República Argentina.
- MORENO, M.; QUINTALES, L.; GARCÍA, F. y MARTÍN P. 2002. "Obtención y Validación de Modelos de Estimación de Software Mediante Técnicas de Minería de Datos". **Revista Colombiana de Computación**. Volumen 3, Número 1. Págs. 53-71. Universidad de Salamanca, Departamento de Informática y Automática. España.
- PADMANABHAN; TUZHILIN. 1999. **Unexpectedness as a measure of Interestingness in Knowledge Discovery System**. Número 27.
- PIATESKY-SHAPIRO, G.; FRAWLEY, W. 1991. **Knowledge Discovery in Databases**. Cambridge, MA:AAA/MIT Press.
- THURASINGHAM, B. 1999. **Data Mining: Technologies, Techniques, Tools and Trends**. CRC Press.