

“MINERIA DE DATOS”

INTRODUCCION:

En la actualidad disponemos de gran cantidad de datos, lo cual hace necesario analizarlos ordenadamente para extraer de un modo automatizado la inteligencia contenida en ellos utilizando técnicas especializadas apoyadas en herramientas informáticas. Estas técnicas constituyen la minería de datos.

Actualmente se utilizan técnicas estadísticas de análisis multivariable de datos de una forma sencilla, al crecer los medios informáticos se ha facilitando de sobremanera la manejabilidad de los algoritmos estadísticos. No obstante, las técnicas de minerías de datos son tan antiguas como la estadística misma.

MINERIA DE DATOS:

“Extracción de patrones(conocimiento) en grandes bases de datos” (Berzal, s.f.)

“Podemos definir la minería de datos como un conjunto de técnicas encaminadas al descubrimiento de la información contenida en grandes conjuntos de datos. Se trata de analizar comportamientos, patrones, tendencias, asociaciones y otras características del conocimiento inmerso en datos.” (Pèrez Marqués, 2014)

“La minería de datos es el proceso que tiene como propósito descubrir, extraer y almacenar información relevante de amplias bases de datos, a través de programas de búsqueda e identificación de patrones y relaciones globales, tendencias, desviaciones y otros indicadores aparentemente caóticos que tienen una explicación que puede descubrirse mediante diversas técnicas de esta herramienta.” (Angles Larrieta & Santillán Gómez)

“La minería de datos consiste en diversos conjuntos de procesos analíticos para explorar grandes conjuntos de datos.” (Caridad y Ocerin)

OBJETIVO DE LA MINERIA DE DATOS:

Aprovechar el valor de la información localizada y usar los patrones preestablecidos para que los directivos tengan un mejor conocimiento de su negocio y puedan tomar decisiones mas confiables.

ESTRUCTURA:

1. Algoritmos o programas de búsqueda mineros: estos programas se utilizan para detectar desviaciones, tendencias y patrones ocultos en los datos históricos. Los mineros son programas pensados y creados por el usuario y su función es correlacionar los criterios de selección y búsqueda con los datos históricos; si encuentran algo interesante lo presentan al usuario como un hallazgo. Los mineros trabajan sobre datos ya recolectados, en máquinas ya existentes, realizando labores útiles mientras los usuarios no se encuentran trabajando.
2. Datos históricos: son datos estables y coherentes que se van acumulando a lo largo de la vida operativa de una empresa.
3. Criterios de búsqueda: son las normas, tendencias desde los cuales los programas mineros realizaran el proceso de selección y búsqueda, los criterios de interés y las explicaciones de situaciones extrañas son definidas por el usuario.
4. Almacenamiento de hallazgos: los hallazgos son datos resultantes de correlacionar los criterios de selección y búsqueda con los datos históricos.

REQUISITOS:

- ✓ No trivial
- ✓ Implícito
- ✓ Previamente desconocido
- ✓ Potencialmente útil

CICLO DE LA MINERIA DE DATOS:

1. Los usuarios de la información deberán identificar los problemas del negocio y las áreas en donde los datos pueden dar valor agregado a la empresa. Asimismo, es importante identificar las áreas en donde la información es muy cambiante, pero primordial para la competitividad de la empresa. Para esto pueden manejarse diversos criterios y no se tiene unos específicos que se les pueda llamar correctos. El objetivo es determinar los criterios, ideas, normas y cuestionamientos que fungirán como entrada para el proceso de minería de datos.
2. Para analizar la información histórica el usuario seleccionara el algoritmo o algoritmos adecuados de minería. Posteriormente, estos algoritmos son traducidos a programas mineros que realizaran las búsquedas con los criterios previamente definidos.
3. Incorporar la información obtenida a través del proceso de minería de datos al proceso de toma de decisiones; así como presentar los hallazgos encontrados a los responsables de las operaciones de forma que la información obtenida pueda integrarse en los procesos de la empresa y pueda aplicarse en la solución de problemas.
4. Medir los resultados: Medir el valor de los hallazgos encontrados, que se proporcionan al tomador de decisiones con relación a la solución de los problemas identificados y a los criterios definidos en el primer punto.

APLICACIONES:

- ✓ Market basket analysis (compras)
- ✓ Ayuda a los directivos a analizar los hábitos de los clientes a fin de satisfacer mejor su demanda.
- ✓ Perfiles de usuario en la Web
- ✓ Análisis diario por producto de la información sobre ventas de un comercio puede incrementar su margen en determinado porcentaje.
- ✓ Segmentación de clientes
- ✓ Para realizar un seguimiento de compras de un cliente en el nivel de canasta de compra o de lo que han comprado en una serie de visitas previas.
- ✓ Detección de fraudes o intrusos
- ✓ La comprensión de los factores que determinan el periodo necesario para liquidar un siniestro para reducir el tiempo medio para tramitarlo.

MODELOS DE MINERIAS DE DATOS:

En función del propósito general:

- ✓ Modelos descriptivos: describen el comportamiento de los datos de forma que sea interpretable por un usuario experto.
- ✓ Modelos predictivos: no solo describen los datos si no que se utilizan para predecir el valor de algún atributo desconocido.

TECNICAS DE MINERÍA DE DATOS:

- ✓ Detección de anomalías (Outliers)
- ✓ Caracterización o resumen
- ✓ Clasificación y predicción
- ✓ Análisis de tendencias (series temporales)
- ✓ Discriminación o contraste
- ✓ Patrones frecuentes, asociaciones y correlaciones
- ✓ Detección de agrupamientos (clustering)

No obstante, las técnicas de minería de datos también se pueden clasificar dependiendo de:

- ✓ El tipo de datos que hay que analizar
- ✓ El tipo de “Conocimiento” que se obtiene
- ✓ El tipo de herramienta que se utiliza
- ✓ El dominio de aplicación

FUENTES DE DATOS:

- ✓ Bases de datos relacionales
- ✓ Bases de datos multidimensionales
- ✓ Bases de datos transaccionales
- ✓ Series temporales, secuencias y data streams
- ✓ Datos estructurados (grafos, redes sociales)
- ✓ Datos espaciales y espaciotemporales
- ✓ Textos e hipertextos (ejemplo la web)
- ✓ Bases de datos multimedia (ejemplo imágenes)

DESCRIPCION DE UNA TAREA DE MINERIA DE DATOS:

1. Datos relevantes: es decir lo que hay que analizar
2. Tipo de conocimiento: es decir lo que se desea obtener
3. Conocimiento previo: para guiar el proceso
4. Medidas de interés: para evaluar los resultados obtenidos
5. Técnicas de representación: para representar los resultados obtenidos.

VENTAJAS SOBRE OTRAS HERRAMIENTAS DE MANEJO DE DATOS:

La minería de datos auxilia a los usuarios empresariales en el procesamiento de reservas de datos para descubrir relaciones de las que, en algunos casos, anteriormente ni siquiera se sospechaba.

Permite la gestión en tiempo real de manera eficaz

Pueden trabajar siguiendo los mismos criterios con grandes cantidades de información histórica.

La minería de datos mediante modelos avanzados y reglas de inducción, puede examinar gran cantidad de datos y encontrar patrones difíciles de identificar a simple vista.

El proceso de búsqueda puede ser realizado por herramientas que automáticamente buscan patrones porque así están programadas y despliegan los tópicos mas importantes.

La información obtenida a través de la minería de datos ayuda a los usuarios a elegir cursos de acción y a definir estrategias competitivas, porque conocen información que sólo ellos pueden emplear.

Aplicable a cualquier tipo de empresa

SOFTWARE DE MINERÍA DE DATOS:

- ✓ KNIME= KNIME <http://www.knime.org/> <http://www.knime.org/>
- ✓ RapidMiner = <http://rapidminer.com/> <http://rapidminer.com/> <http://rapidminer.com/>
<http://rapidminer.com/>
- ✓ Weka = <http://www.cs.waikato.ac.nz/ml/weka/>

- ✓ R = <http://www.r-project.org/>
- ✓ SPSS Modeler = <http://www.spss.com/software/modeler/>
- ✓ SAS Enterprise Miner = <http://www.sas.com>

CONCLUSIONES:

En la actualidad la información ha incrementado de sobremanera su valor hasta llegar a convertirse en un activo estratégico para la competitividad de una empresa. La minería de datos ayuda a los directivos a obtener una visión completa y detallada de su negocio ya que les permite buscar datos que salen de ciertos rangos, y no obstante esta información es confiable para toma de decisiones.

La minería de los datos es el futuro puesto que cada vez las empresas se ahogan en un mar de información y se les dificulta obtener solo cierta información en medio de este mar.

AGRADECIMIENTOS:

Le agradezco a mi alma mater el Instituto Tecnológico de Orizaba, al profesor Fernando Aguirre y Hernández quien imparte la materia de Fundamentos de la Ingeniería Administrativa por demostrarnos que somos capaces de escribir artículos de diversos temas, por fomentarnos el hábito de la lectura y sobre todo por ayudarnos a darnos cuenta de lo que somos capaces de lograr.

BIBLIOGRAFÍA

Angles Larrieta, M., & Santillán Gómez, A. (s.f.). Minería de datos: Concepto, características, estructura y aplicaciones. *CA*, 79-84.

Berzal, F. (s.f.). *DECSAI*. Recuperado el MARZO de 2016, de Departamento de Ciencias de la computacion e L.A:
<http://elvex.ugr.es/decsai/intelligent/slides/dm/D1%20Data%20Mining.pdf>

Pèrez Marqués, M. (2014). *MINERÍA DE DATOS A TRAVÉS DE EJEMPLOS*. ESPAÑA: RC LIBROS.