



Unidad # 3

Almacenes de Datos y Minería de Datos

Administración de Bases de Datos II

Gisela Yasmín García Espinoza



**UNIVERSIDAD
GERARDO BARRIOS**
Líderes en Gestión del Conocimiento





Competencia de la asignatura

Crear bases de datos no relacionales y consistentes para almacenar la data empresarial aplicando la metodología del diseño, gestión de base de datos, lenguaje No SQL y considerando las reglas de negocios desarrolladas de manera individual y colaborativa.

Contenidos a desarrollar

- Introducción a la minería de datos
- Técnicas de la minería de datos



Data Mining



Introducción

Con los rápidos avances en la tecnología de la información, se observa un crecimiento explosivo en la generación de datos y las capacidades de recopilación de datos en todos los dominios.

En el mundo empresarial, los minoristas y el comercio electrónico han generado grandes bases de datos sobre transacciones comerciales. También se han generado una gran cantidad de datos científicos en varios campos. La World Wide Web proporciona un ejemplo con miles de millones de páginas web que consisten en información textual y multimedia que son utilizadas por millones de personas.

El análisis de grandes cantidades de datos que se pueden comprender y utilizar de manera eficiente sigue siendo un problema desafiante. La minería de datos aborda este problema proporcionando técnicas y software para automatizar el análisis y la exploración de conjuntos de datos grandes y complejos. La investigación sobre minería de datos se lleva a cabo en una amplia variedad de campos, que incluyen estadísticas, informática, aprendizaje automático, gestión de bases de datos y visualización de datos, por nombrar algunos.

Minería de Datos

La minería de datos , también conocida popularmente como descubrimiento de conocimientos en bases de datos (KDD), se refiere a la extracción no trivial de información implícita, previamente desconocida y potencialmente útil de datos en bases de datos. Si bien la minería de datos y el descubrimiento de conocimientos en bases de datos (o KDD) se tratan con frecuencia como sinónimos, la minería de datos es en realidad parte del proceso de descubrimiento de conocimientos.

“ Data mining is the practice of searching through large amounts of computerized data to find useful patterns or trends.

Merriam-Webster Dictionary

SMARTER**CX**



Fuente: https://smartercx.com/wp-content/uploads/2019/02/041019_SmarterCX_Data_Mining_1024x512_01.jpeg

Minería de Datos

La minería de datos se refiere a un conjunto de métodos aplicables a bases de datos grandes y complejas para eliminar la aleatoriedad y descubrir el patrón oculto. Los métodos de minería de datos casi siempre son computacionalmente intensivos.

La minería de datos se trata de herramientas, metodologías y teorías para revelar patrones en los datos, que es un paso fundamental en el descubrimiento del conocimiento.

Minería de Datos

Utilizamos técnicas de minería de datos para un largo proceso de investigación y desarrollo de productos. Como esta evolución se inició cuando los datos comerciales se almacenaron por primera vez en las computadoras. Además, permite a los usuarios navegar por sus datos en tiempo real. La minería de datos en la comunidad empresarial está respaldada por tres tecnologías que ahora están maduras:

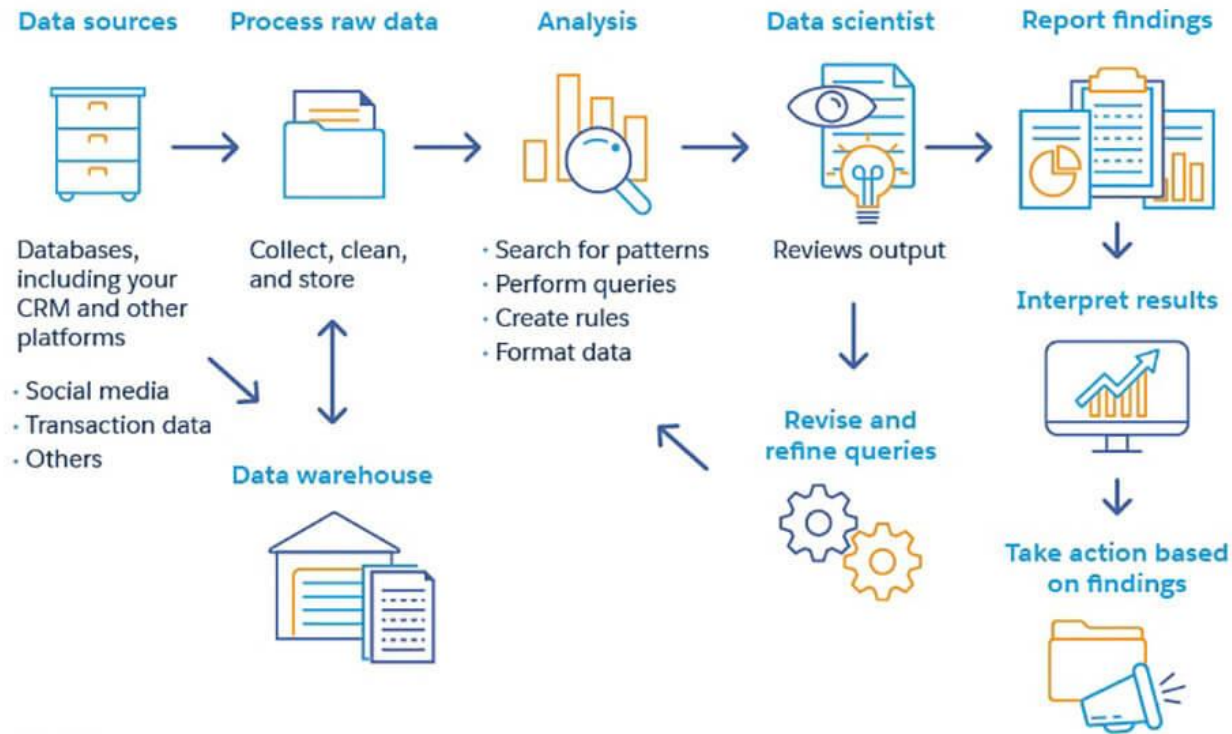
Recopilación masiva de datos

Potentes ordenadores multiprocesador

Algoritmos de minería de datos



How Data Mining Works



SOURCE:

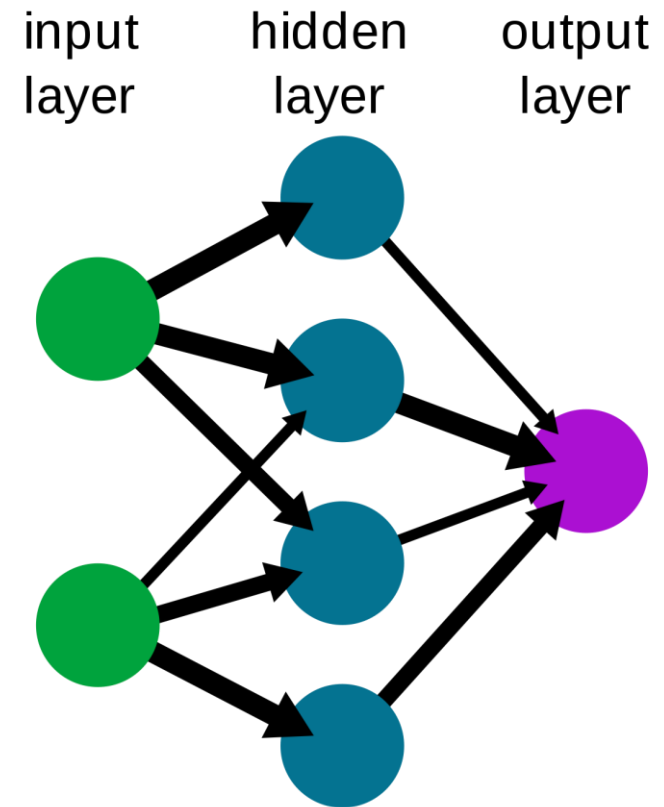
[slideshare.net/PowerPoint-Templates/data-mining-process-powerpoint-presentation-templates](https://www.slideshare.net/PowerPoint-Templates/data-mining-process-powerpoint-presentation-templates)

Técnicas de Minería de Datos

Redes neuronales artificiales

Las redes neuronales son un medio de hacer aprendizaje automático, en el que una computadora aprende a realizar alguna tarea mediante el análisis de ejemplos de entrenamiento. Por lo general, los ejemplos se han etiquetado a mano con anticipación. Un sistema de reconocimiento de objetos, por ejemplo, podría alimentar miles de imágenes etiquetadas de automóviles, casas, tazas de café, etc., y encontraría patrones visuales en las imágenes que se correlacionan consistentemente con etiquetas particulares.

A simple neural network



Fuente: https://upload.wikimedia.org/wikipedia/commons/thumb/9/99/Neural_network_example.svg/1200px-Neural_network_example.svg.png

Usos de la Minería de Datos

Predicción automatizada de tendencias y comportamientos.

Utilizado para automatizar el proceso de búsqueda de información predictiva en grandes bases de datos. Las preguntas que requerían un análisis práctico extenso ahora se pueden responder a partir de los datos. El marketing dirigido es un ejemplo típico de marketing predictivo. Como también utilizamos la minería de datos en correos promocionales anteriores. Eso es para identificar los objetivos para maximizar el retorno de la inversión en futuros envíos. Otros problemas de predicción incluyen la previsión de quiebras y otras formas de incumplimiento. E identificar segmentos de una población que probablemente respondan de manera similar a eventos dados.

Usos de la Minería de Datos

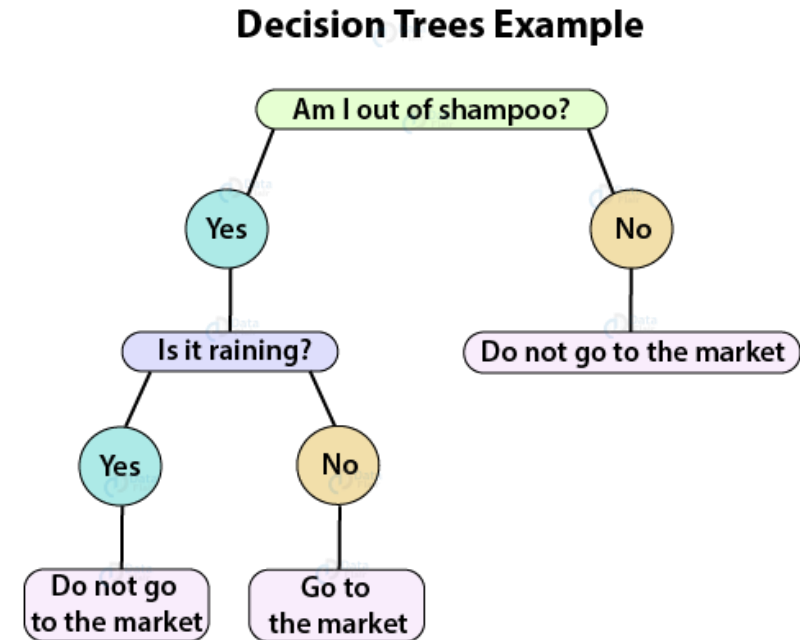
Descubrimiento automatizado de patrones previamente desconocidos.

Hay un muy buen ejemplo de descubrimiento de patrones. Como es el análisis de datos de ventas minoristas. Eso es para identificar productos no relacionados que a menudo se compran juntos. Además, existen otros problemas de descubrimiento de patrones. Eso incluye la detección de transacciones fraudulentas con tarjetas de crédito. Se identifica que los datos anómalos podrían representar errores de introducción de datos.

Técnicas de Minería de Datos

Arboles de decisión

Un árbol de decisiones es un diagrama o gráfico que la gente usa para determinar un curso de acción o mostrar una probabilidad estadística. Forma el contorno de la planta leñosa homónima, generalmente erguida pero a veces acostada de lado. Cada rama del árbol de decisiones representa una posible decisión, resultado o reacción. Las ramas más lejanas del árbol representan los resultados finales.



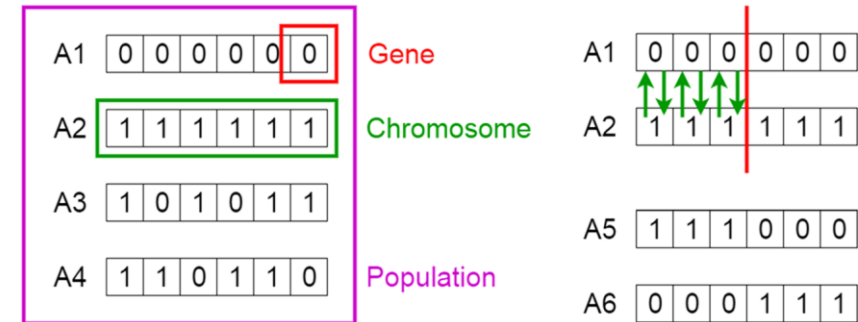
Fuente: <https://data-flair.training/blogs/wp-content/uploads/sites/2/2017/07/Decision-Trees-Example.png>

Técnicas de Minería de Datos

Algoritmos genéticos

Los algoritmos genéticos (GA) son algoritmos basados en búsquedas que se basan en los conceptos de selección natural y genética. Los GA son un subconjunto de una rama mucho más grande de la computación conocida como Computación Evolutiva .

Genetic Algorithms



Fuente: https://miro.medium.com/max/1838/1*BYDJpa6M2rzWNSurvspf8Q.png

Sitios consultados

<https://medium.com/@patelharshali136/introduction-to-data-mining-and-machine-learning-techniques-1d209b6c69b2>

<https://webdocs.cs.ualberta.ca/~zaiane/courses/cmp690/notes/Chapter1/index.html>

Tabla de actividades

Nombre de la actividad	Técnicas de minería de datos
Tipo de actividad	Tarea
Tipo de participación	Individual
Competencias específica de la asignatura	Crear bases de datos no relacionales y consistentes para almacenar la data empresarial aplicando la metodología del diseño, gestión de base de datos, lenguaje No SQL y considerando las reglas de negocios desarrolladas de manera individual y colaborativa.
Instrucciones	Después de haber dado lectura a los contenidos de la semana y participar activamente en la video conferencia, deberá seleccionar una técnica de minería de datos y responder en el foro correspondiente en la semana, descripción de la técnica (200 palabras) y un ejemplo de uso.
Fecha de entrega	La fecha límite de participación será el día de desarrollo de la clase
Instrumento de evaluación	<ul style="list-style-type: none"> - Puntualidad 20% - Coherencia 20% - Contenido(agrega referencias, respeta cantidad de palabras) 40% - Ejemplo 20%
Ponderación	Coevaluación: 30% Laboratorio II

Recursos Complementarios

Recurso	Título	Cita Referencial
Sitio Web	Introducción a las técnicas de minería de datos	https://medium.com/@patelharshali136/introduction-to-data-mining-and-machine-learning-techniques-1d209b6c69b2
Sitio Web	¿Qué es la minería de datos?	https://online.stat.psu.edu/stat857/node/142/

¿Preguntas?

¡Muchas gracias!