



# MINERÍA DE DATOS

Padilla Martin Juan Pablo 217294261

Álvarez Gutiérrez David Alejandro 217294016

Ruiz Ortiz Cesar Adrián 208020817



# EL POR QUE DE LA MINERIA DE DATOS

- Actualmente todas las ramas, tanto de estudio y académicas, están generando grandes cantidades de petabytes por lo cual es importante ver que de esa información que se genera continuamente es valiosa para distintas ramas y transformar esa información en conocimiento organizado.

# EJEMPLOS DE USO DE LA MINERIA

- **Motores de búsqueda:** Actualmente por medio de motores de búsqueda, como Google, se puede detectar información valiosa, como que tanta gente busca información de enfermedades y así poder anticipar problemas grandes de salud.

- **Marketing:** La minería de datos se utiliza para explorar bases de datos cada vez mayores y mejorar la segmentación del mercado. Analizando las relaciones entre parámetros como edad de los clientes, género, gustos, etc.

**Comercio minorista:** Los supermercados, por ejemplo, emplean los patrones de compra conjunta para identificar asociaciones de productos y decidir cómo situarlos en los diferentes pasillos y estanterías de los lineales.

- **Banca:** Los bancos recurren a la minería de datos para entender mejor los riesgos del mercado. Es habitual que se aplique a la calificación crediticia (rating) y a sistemas inteligentes antifraude para analizar transacciones, movimientos de tarjetas, patrones de compra y datos financieros de los clientes.

# EVOLUCION DE LA MINERIA DE DATOS

- 1960 :  
Recopilación de datos y creación de bases de datos
- 1970: Sistemas de gestión de bases de datos, Sistemas de bases de datos jerárquicos y de red. Sistemas de bases de datos relacionales Modelado de datos: modelos de entidad-relación, etc. Métodos de indexación y acceso Lenguajes de consulta: SQL, etc. Interfaces de usuario, formularios e informes. Consulta de procesamiento y optimización Transacciones, control de concurrencia y recuperación. Procesamiento de transacciones en línea (OLTP)

- 1980: Sistemas avanzados de bases de datos, Modelos de datos avanzados: relacional extendido, objeto relacional, deductivo, etc. Gestión de datos complejos, Flujos de datos y sistemas de datos ciberfísicos, Gestión de datos extremadamente grande. ETC
- FINALES DE LOS 80: Almacén de datos y OLAP Minería de datos y descubrimiento de conocimiento: clasificación, agrupamiento, análisis de valores atípicos, asociación y correlación, comparativa resumen, análisis de discriminación, patrón descubrimiento, análisis de tendencias y desviaciones, etc.
- PRESENTE: Generación futura de sistemas de información



# ¿Qué es la minería de datos?

- minería de conocimiento a partir de datos, extracción de conocimiento, análisis de datos / patrones, datos arqueología y dragado de datos.

# PASOS PARA LA OBTENCION DE INFORMACION:

- 1. Limpieza de datos (para eliminar ruido y datos inconsistentes).
- 2. Integración de datos (donde se pueden combinar múltiples fuentes de datos).
- 3. Selección de datos (donde los datos relevantes para la tarea de análisis se recuperan de base de datos).
- 4. Transformación de datos (donde los datos se transforman y consolidan en formularios apropiado para la minería realizando operaciones de resumen o agregación).

- 5. Minería de datos (un proceso esencial donde se aplican métodos inteligentes para extraer patrones de datos).
- 6. Evaluación de patrones (para identificar los patrones verdaderamente interesantes que representan el conocimiento basado en medidas de interés).
- 7. Presentación del conocimiento (donde las técnicas de visualización y representación del conocimiento se utilizan para presentar el conocimiento minado a los usuarios).

# ¿Qué tipos de datos se pueden extraer?

- Las formas más básicas de datos para la minería.
- datos de la base de datos.
- datos del almacén de datos.
- datos transaccionales.

# ¿Qué tipos de patrones se pueden extraer?

- caracterización y discriminación.
- la minería de patrones frecuentes, asociaciones y correlaciones.
- clasificación y regresión.
- análisis de agrupamiento.
- y análisis de valores atípicos.

# ¿Qué tecnologías se utilizan?

- **La estadística** estudia la recopilación, análisis, interpretación o explicación y presentación de datos.
- **El aprendizaje automático (MACHINE LEARNING)** investiga cómo las computadoras pueden aprender (o mejorar su rendimiento) basado en datos.

- **La recuperación de información (IR)** es la ciencia de buscar documentos o información en documentos. Los documentos pueden ser de texto o multimedia, y pueden residir en la Web.
- **La investigación de sistemas de bases de datos** se centra en la creación, mantenimiento y uso de bases de datos. para organizaciones y usuarios finales.

# ¿Qué tipos de aplicaciones están dirigidas?

- **Inteligencia de negocios:** Es fundamental que las empresas adquieran una mejor comprensión del contexto comercial. de su organización, como sus clientes, el mercado, la oferta y los recursos, y competidores.
- **motor de búsqueda web:** es un servidor informático especializado que busca información En la red. Los resultados de búsqueda de una consulta de usuario a menudo se devuelven como una lista.



# Mapa conceptual

