

Minería de datos

Tecnológico de Costa Rica

Programa de Ciencia de Datos

Frans van Dunné

Agenda

- 8:00 – 9:30
 - Fundamentos de Inteligencia de negocios.
 - Concepto de inteligencia de negocio.
 - Tipos de datos explotados.
- 9:20 – 9:35
 - Pausa
- 9:35 – 12:00
 - Recogida y preparación de datos.
 - Metodologías (KDD, CRISP-DM)
 - Casos de Estudio



“ Inteligencia de Negocio se refiere al proceso de convertir datos en conocimiento y conocimiento en acciones para crear la ventaja competitiva del negocio “

The Data Warehousing Institute

Datos

- Los datos son la mínima unidad semántica, y se corresponden con elementos primarios de información que por sí solos son irrelevantes como apoyo a la toma de decisiones. También se pueden ver como un conjunto discreto de valores, que no dicen nada sobre el por qué de las cosas y no son orientativos para la acción.

Información

- La información se puede definir como un conjunto de datos procesados y que tienen un significado (relevancia, propósito y contexto), y que por lo tanto son de utilidad para quién debe tomar decisiones, al disminuir su incertidumbre. Los datos se pueden transformar en información añadiéndoles valor:
 - **Contextualizando:** se sabe en qué contexto y para qué propósito se generaron.
 - **Categorizando:** se conocen las unidades de medida que ayudan a interpretarlos.
 - **Calculando:** los datos pueden haber sido procesados matemática o estadísticamente.
 - **Corrigiendo:** se han eliminado errores e inconsistencias de los datos.
 - **Condensando:** los datos se han podido resumir de forma más concisa (agregación).

Conocimiento

- El conocimiento es una mezcla de experiencia, valores, información, que sirve como marco para la incorporación de nuevas experiencias e información, y es útil para la acción.
- Se origina y aplica en la mente de los conocedores. En las organizaciones con frecuencia no sólo se encuentra dentro de documentos o almacenes de datos, sino que también está en rutinas organizativas, procesos, prácticas, y normas.

Conocimiento

- El conocimiento se deriva de la información, así como la información se deriva de los datos. Para que la información se convierta en conocimiento es necesario realizar acciones como:
 - Comparación con otros elementos.
 - Predicción de consecuencias.
 - Búsqueda de conexiones.
 - Conversación con otros portadores de conocimiento.

Inteligencia de Negocios **TEC** | Tecnológico de Costa Rica

- Inteligencia de negocios o **BI** (del inglés business intelligence) es el conjunto de estrategias y herramientas para la administración y creación de conocimiento mediante el análisis de la información de una organización para facilitar la toma de decisiones empresariales, predicciones y perfiles futuras.

Fundamentos de Inteligencia de negocios

- La utilización de los datos en las organizaciones como elemento que facilita la toma de decisiones, involucra el conocimiento del funcionamiento actual y la anticipación de acontecimientos futuros.
- Alrededor de este tema hay toda una serie de estrategias y herramientas las cuales a partir del análisis de datos existentes, generan conocimiento que permite apoyar las decisiones empresariales.

Fundamentos de Inteligencia de negocios

Ejercicio (10 min):

Presenten la lista de estrategias y herramientas de BI que conocen, las cuales a partir del análisis de datos existentes, generan conocimiento que permite apoyar las decisiones empresariales

¿Qué hace la inteligencia de negocios?

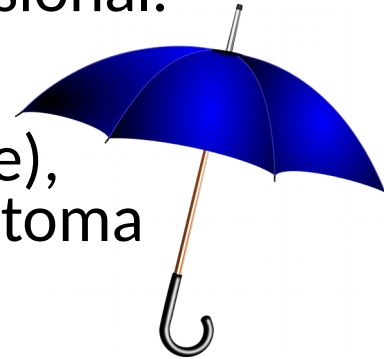
- Genera reportes globales o por secciones.
- Crea escenarios con respecto a una decisión.
- Hace pronósticos de ventas y devoluciones.
- Comparte información entre departamentos.
- Hace análisis multidimensionales.
- Genera y procesa datos.
- Cambia la estructura de toma de decisiones.
- Mejora el servicio al cliente.

Fundamentos de Inteligencia de negocios

- En 1958 Hans Peter Luhn habla por primera vez sobre inteligencia de negocios, entendida como la interrelación de hechos orientados al desarrollo de una meta.
- Anteriormente las empresas generaban datos y con mucho esfuerzo generaban conocimientos sobre estos, debido a que no contaban con recursos computacionales, dificultando el análisis de los datos por la carencia de herramientas que permitieran la toma de decisiones a largo plazo y las decisiones a corto plazo se toman por intuición.

Fundamentos de Inteligencia de negocios

- La inteligencia de negocios permite la toma de decisión en los niveles estratégicos, tácticos u operativos, extrayendo datos para analizarlos, generar reportes y realizar análisis cruzados por medio del análisis multidimensional.
- Howard Dresner en 1985, puso el acrónimo de BI (Business Intelligence), como el conjunto de métodos, herramientas y procedimientos para la toma de decisiones, basados en hechos.
- En la actualidad BI, incluye una amplia gama de metodologías, aplicaciones y tecnologías, que permiten extraer de diversas fuentes, plataformas y estructuras para transformarlas y analizarlas: datos, transacciones e información son convertidas en conocimientos para soporte en la toma de decisiones.



Fundamentos de Inteligencia de negocios

- La Inteligencia de Negocios se centra en el manejo de grandes volúmenes de información provenientes de las diferentes áreas de la empresa, a esta información se le aplican metodologías y herramientas para que sea posible analizar el entorno organizacional determinando fortalezas y debilidades para así establecer estrategias y el descubrimiento de nuevas oportunidades.

La implementación de un esquema de BI permite:

- Información correcta en el momento adecuado, almacenada en un único lugar y disponible en tiempo real.
- Evaluación de distintos escenarios que permiten analizar situaciones que puedan afectar al negocio.
- La información va mas allá de los reportes, incluyendo indicadores que me permitan medir el desempeño del negocio.

Factores de éxito de la inteligencia de negocios



- Proveen acceso a datos adecuados.
- Incrementan la habilidad de los usuarios para entender los resultados.
- Incrementan el entendimiento de los negocios por parte de los usuarios.
- Ayudan a comunicar los hallazgos y tomar acciones.

Factores de éxito de la inteligencia de negocios

TEC | Tecnológico
de Costa Rica

Ejercicio (10 min)

Cual de estos es el mayor reto en la implementacion de BI:

- Proveen acceso a datos adecuados.
- Incrementan la habilidad de los usuarios para entender los resultados.
- Incrementan el entendimiento de los negocios por parte de los usuarios.
- Ayudan a comunicar los hallazgos y tomar acciones.

Beneficios de la inteligencia de negocios



- Toman mejores decisiones con una asombrosa velocidad y confianza.
- Dinamizan sus operaciones .
- Reducen los ciclos de vida de sus productos.
- Maximizan el valor de las líneas de producto y anticipan nuevas oportunidades.
- Hacer un mejor y más enfocado marketing mejorando las relaciones con los clientes y proveedores por iguales.

¿Por qué falla la inteligencia de negocios?

- Por la falta de compromisos en las autoridades de la empresa.
- Se tiene poca disponibilidad de los representantes de negocios.
- Hay ausencia de un personal disponible y habilidoso.
- Existe un mal concepto del software de BI.
- No trabajan bajo una estructura detallada.
- No existe un análisis del negocio o estandarización.
- No existe una apreciación del impacto que causan los datos de mala calidad en la rentabilidad del negocio.
- No se entiende la necesidad del uso de un meta datos.
- Demasiada confianza métodos y herramientas no alineadas.

Riesgos de la inteligencia de negocios

- La tecnología esta cambiando rápidamente.
 - Los datos que son usados no son transformados apropiadamente.
 - La selección del proveedor más apropiados y sus productos, así como su implementación.
- Genera reportes globales o por secciones.
 - Crea escenarios con respecto a una decisión.
 - Hace pronósticos de ventas y devoluciones.
 - Comparte información entre departamentos.
 - Hace análisis multidimensionales.
 - Genera y procesa datos.
 - Cambia la estructura de toma de decisiones.
 - Mejora el servicio al cliente.

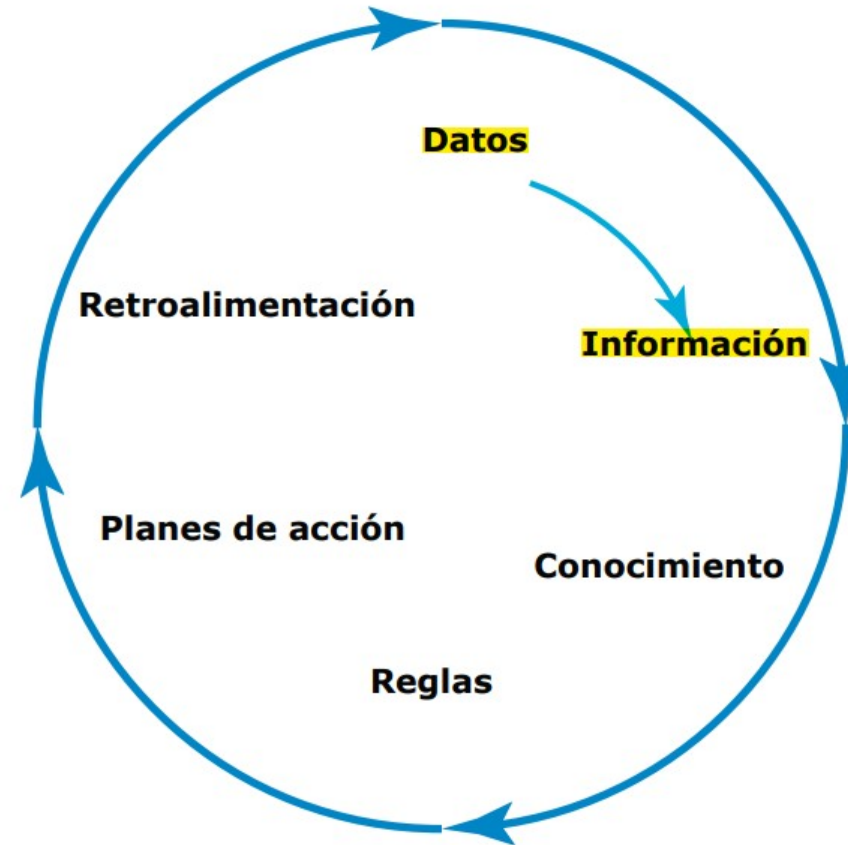
La implementación de un esquema de BI permite:

- Permite agrupar información de diferentes áreas para establecer, modificar y ajustar políticas, procesos y procedimientos.
- Gran capacidad de reacción ante amenaza y vulnerabilidades del negocio, actuando o anticipándose a éstos.
- Capacidad de retroalimentación con respecto a hechos sucedidos para reevaluar políticas y objetivos del negocio.

Las 4 P de BI:

- **PLAN:** curso de acción conscientemente determinado.
- **POSICIÓN:** un medio para ubicar a la organización (nicho, rentas, dominio) .
- **PATRÓN:** es un modelo que implica consistencia.
- **PERSPECTIVA:** una manera particular de percibir el mundo (concepto, cultura, ideología) .

La implementación de un esquema de BI permite:



Selección de Indicadores **TEC** | Tecnológico de Costa Rica

- Para generar el concepto de BI es importante entender cuál es el flujo de información en la empresa y cuáles son los procesos involucrados en el requerimiento de información.
- Por eso es necesario preguntarse:
 - ¿Qué información se necesita?
 - ¿Para qué se quiere la información?
 - ¿A quien va dirigida la información?

Selección de Indicadores **TEC** | Tecnológico de Costa Rica



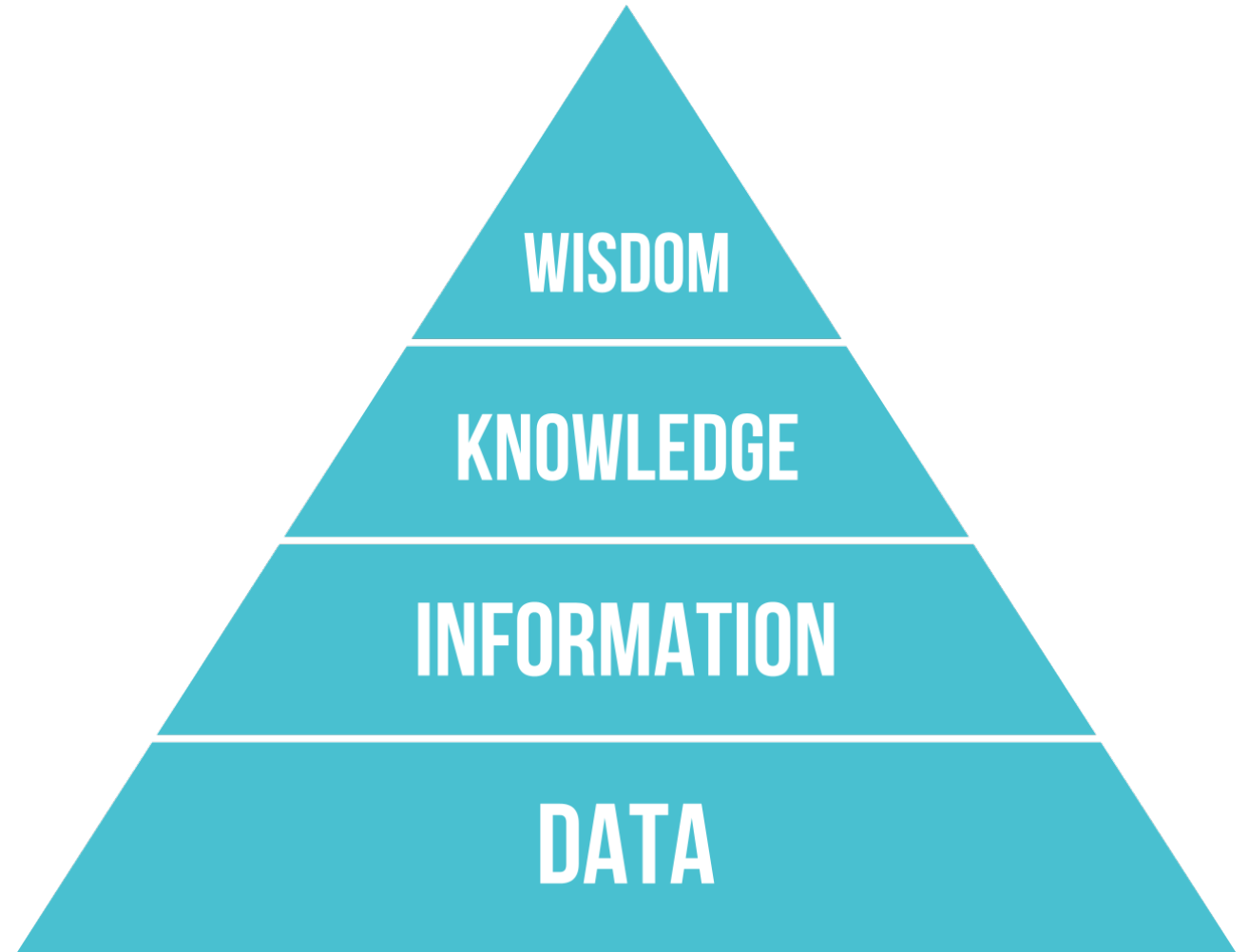
Selección de Indicadores **TEC** | Tecnológico de Costa Rica



Ejercicio (15 min)

El profe quiere implementar BI para mejorar sus clases. Cuales serian ejemplos a estos cuatro niveles en ese caso?

Selección de Indicadores **TEC** | Tecnológico de Costa Rica



Definición de componentes claves de un sistema BI

- **Fuentes de datos.** Los de origen interno provienen de *sistemas Customer Relationship Management (CRM)*, *sistemas Enterprise Resource Planning (ERP)* y bases de datos o software de inventario, RR.HH., contabilidad, etc. Por otra parte, los de origen externo como los datos obtenidos de redes sociales (*opendata*).
- **Procesos ETL.** Se refiere a tres acciones. Extracción de los datos de fuentes. Transformación de esos datos en un formato homogéneo para relacionarlos unos con otros. Cargar los datos en un *Data Warehouse* para después analizarlos.

Definición de componentes claves de un sistema BI

- **Data Warehouse.** Llamado también almacén de datos, es una base de datos que se compone de información procedente de numerosas fuentes que han pasado por un proceso ETL para integrarse de manera homogénea. Está diseñada para organizar y optimizar los datos para los posteriores análisis complejos.
- **Online Analytical Processing (OLAP).** Mientras que los procesos ETL y *Data Warehouse* representan el *back-end* de un sistema de BI, el OLAP representa el *front-end*. En otras palabras, los primeros corresponden al procesamiento esencial de datos. El segundo ofrece una forma completa de visualizar los datos, que también es compleja.
- **Herramientas de visualización.** Es la parte más visible del sistema de BI. Allí suele encontrarse el Cuadro de Mando o *Dashboard*. Suele ser un resumen de los indicadores más relevantes de cara al usuario del sistema. Permite acceder a un cuadro de mando integral.

Tipos de datos explotados

- “Los sistemas transaccionales son los sistemas enfocados en la toma de decisiones, que tienen como función la manipulación de la información, con el fin de apoyar y fundamentar la toma de decisiones
- Dentro de los procesos transaccionales podemos encontrar el ERP, el SCM, y el CRM.

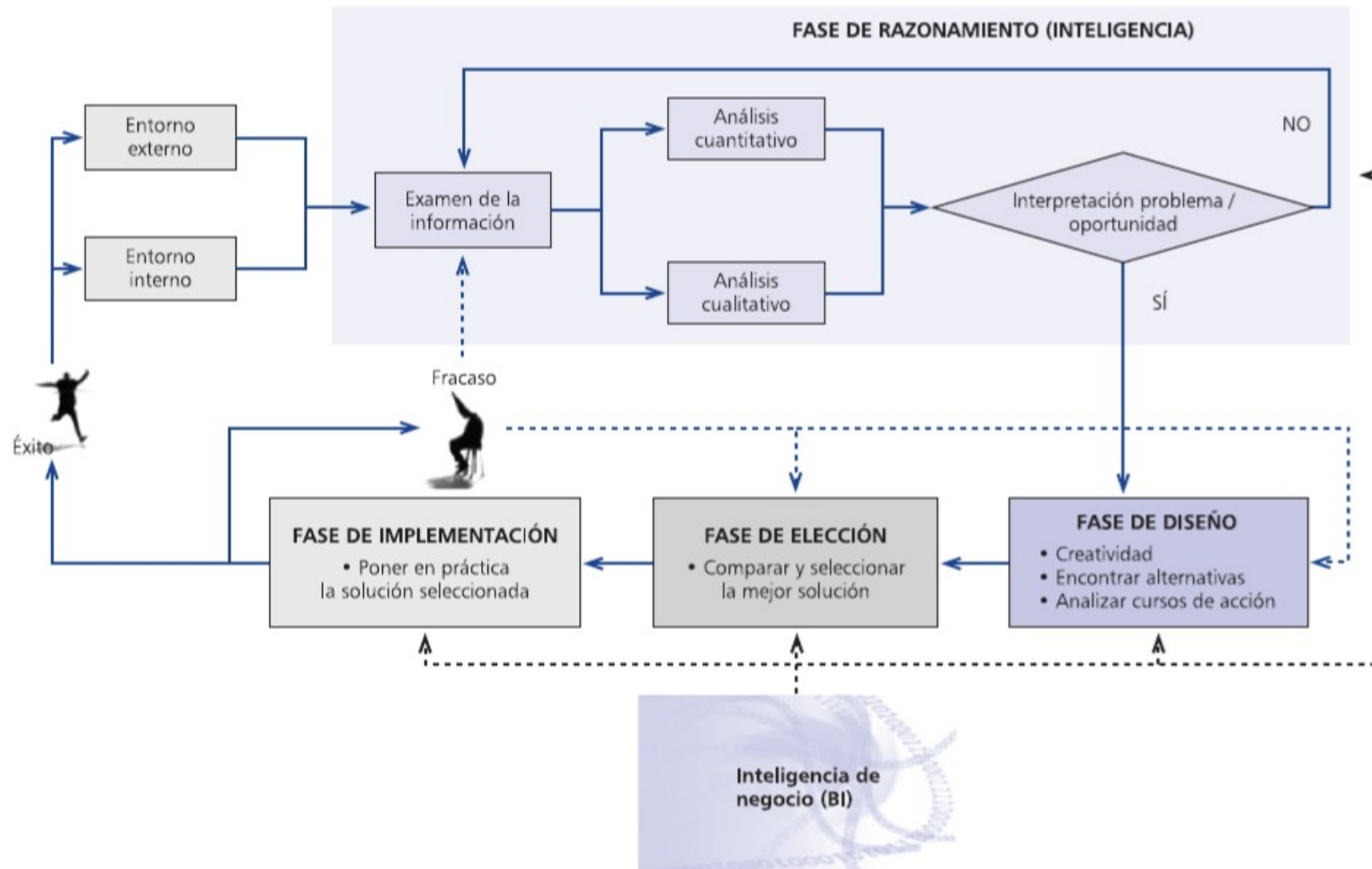
Tipos de datos explotados

- Este sistema de negocio puede contener varios módulos puesto comparten la misma base de datos; es así: **la gestión de producción, gestión de clientes, compras, cuentas a pagar y cobrar, contabilidad general, facturación, gestión de inventario, recursos humanos, nóminas o cualquier otra función que se tenga que desarrollar dentro de la empresa.**



Recogida y preparación de datos

Recogida y preparación de datos.



Recogida y preparación de datos

- **Departamento de Marketing**

- El BI permite identificar de forma más precisa los segmentos de clientes y estudiar con mayor detalle su comportamiento. Para ello se pueden incluir análisis capaces de medir, por ejemplo, el impacto de los precios y las promociones en cada segmento.

- **Departamento de Compras**

- El BI permite acceder a los datos del mercado, vinculándolos con la información básica necesaria para hallar las relaciones entre coste y beneficio. Al mismo tiempo, permite monitorizar la información de cada factoría o cadena de producción, lo que puede ayudar a optimizar el volumen de las compras.

- **Departamento de Producción**

- El BI proporciona un mecanismo que permite analizar el rendimiento de cualquier tipo de proceso operativo, ya que comprende desde el control de calidad y la administración de inventarios hasta la planificación y la historización de la producción.

- **Departamento de Ventas**

- El BI facilita la comprensión de las necesidades del cliente, así como responder a las nuevas oportunidades del mercado. También son posibles análisis de patrones de compra para aprovechar coyunturas de ventas con productos asociados.

Recogida y preparación de datos

- **Departamento Económico-Financiero**

- El BI permite acceder a los datos de forma inmediata y en tiempo real, mejorando así ciertas operaciones, que suelen incluir presupuestos, proyecciones, control de gestión, tesorería, balances y cuentas de resultados.

- **Departamento de Atención al Cliente**

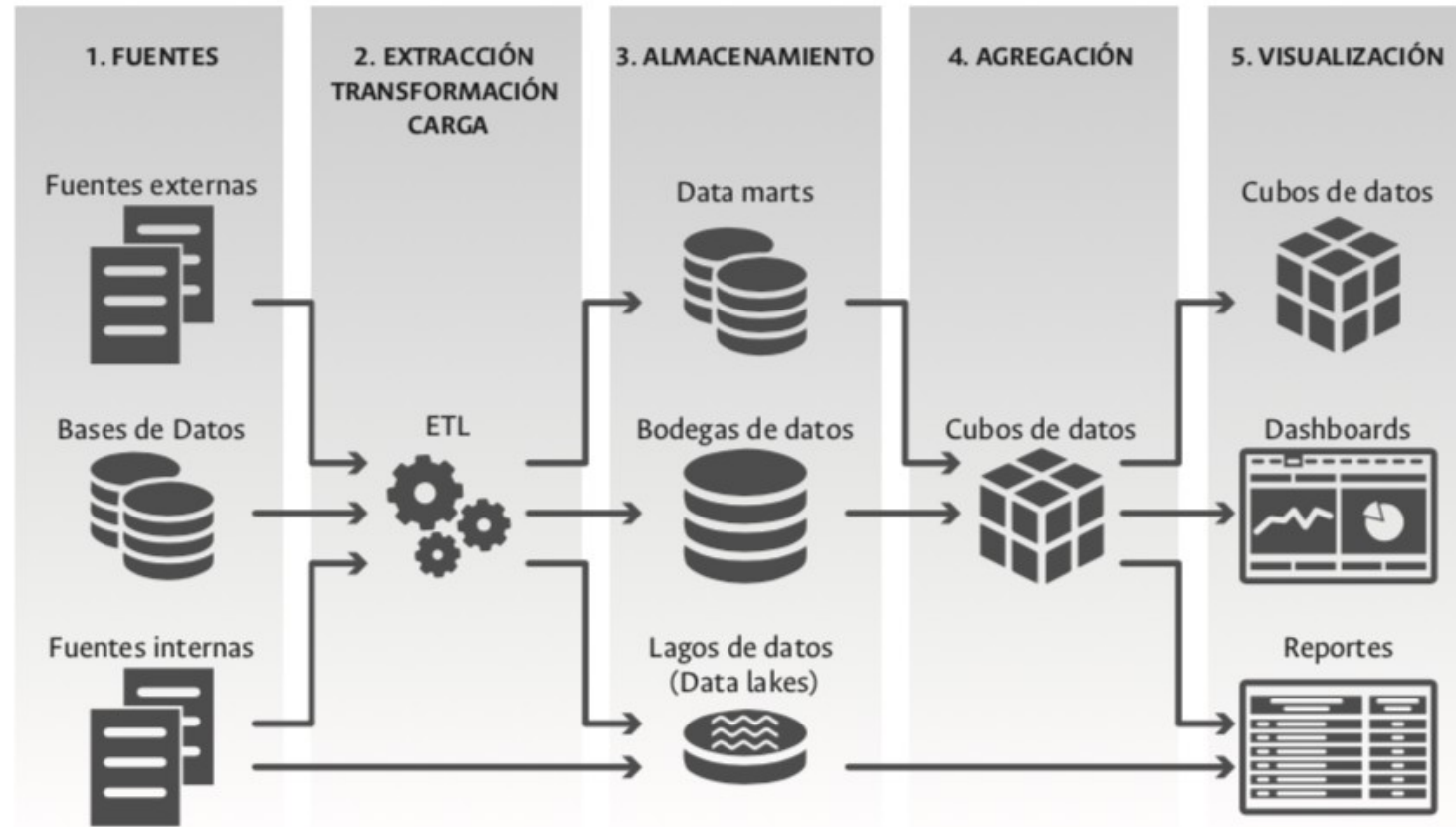
- Aplicado a este ámbito, el BI permite evaluar con exactitud el valor de los segmentos del mercado y de los clientes individuales, además de ayudar a retener a los clientes más rentables.

- **Departamento de Recursos Humanos**

- Obteniendo los datos precisos de la fuente adecuada, el BI permite analizar los parámetros que más pueden afectar al departamento: satisfacción de los empleados, absentismo laboral, beneficio-hora/hombre... etc.

Recogida y preparación de datos

TEC | Tecnológico de Costa Rica

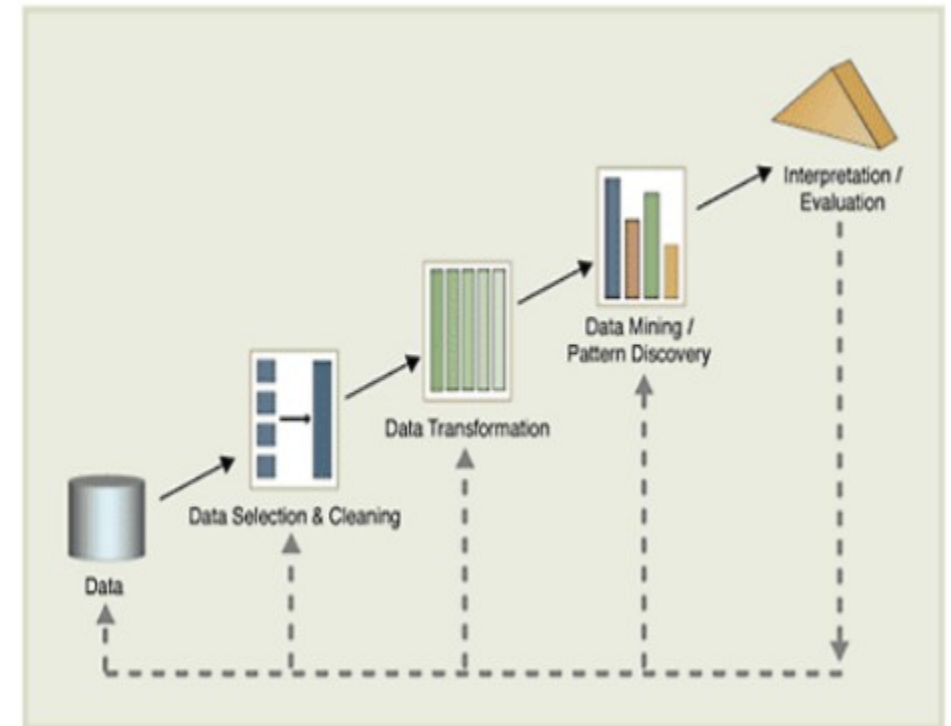


Metodologías

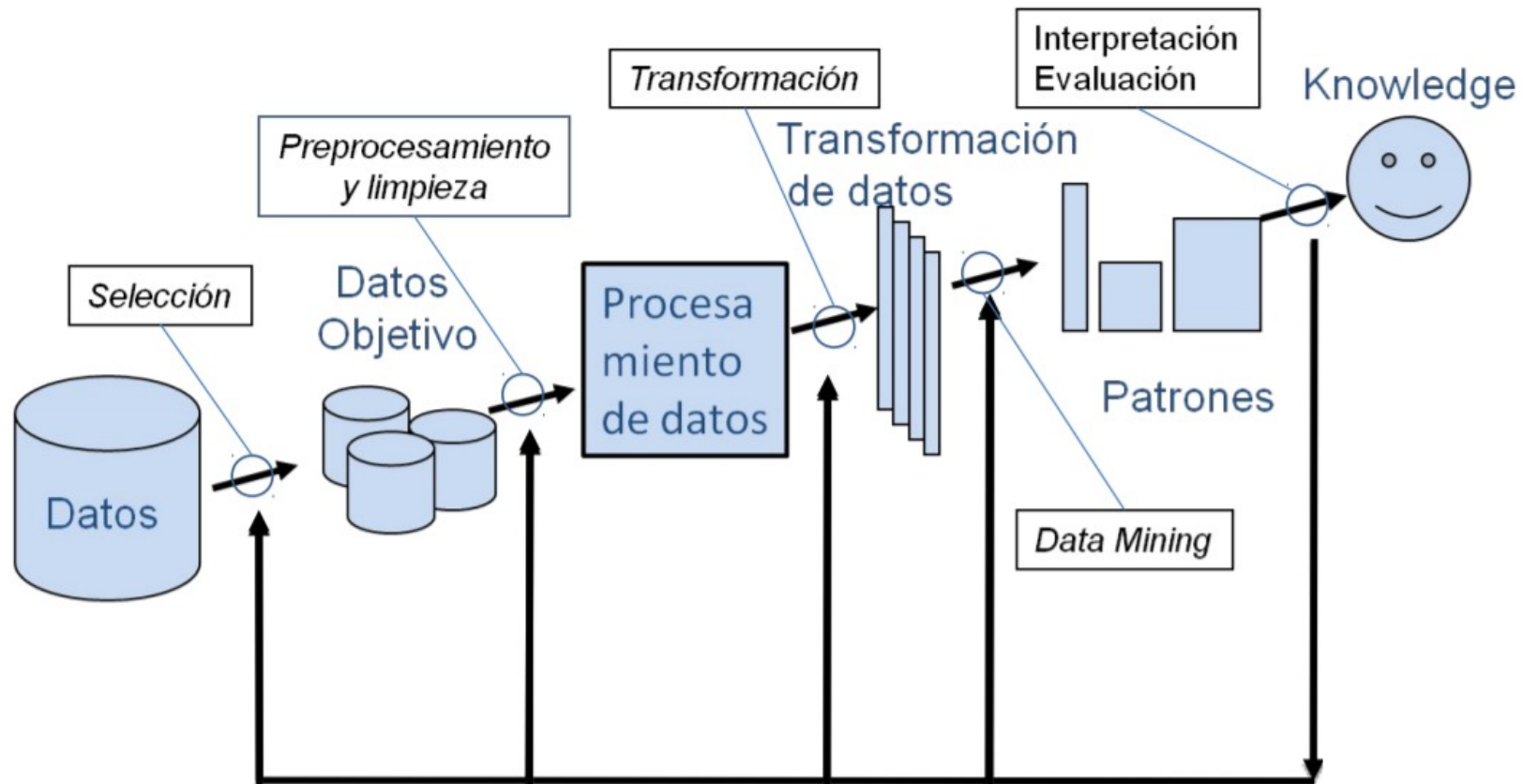
- Las tres metodologías dominantes para el proceso de la minería de datos son: **KDD, CRISP-DM y SEMMA**

KDD

- Es una metodología propuesta por Fayyad en 1996, propone 5 fases: **Selección, preprocesamiento, transformación, minería de datos y evaluación e implantación**. Es un proceso iterativo e interactivo.



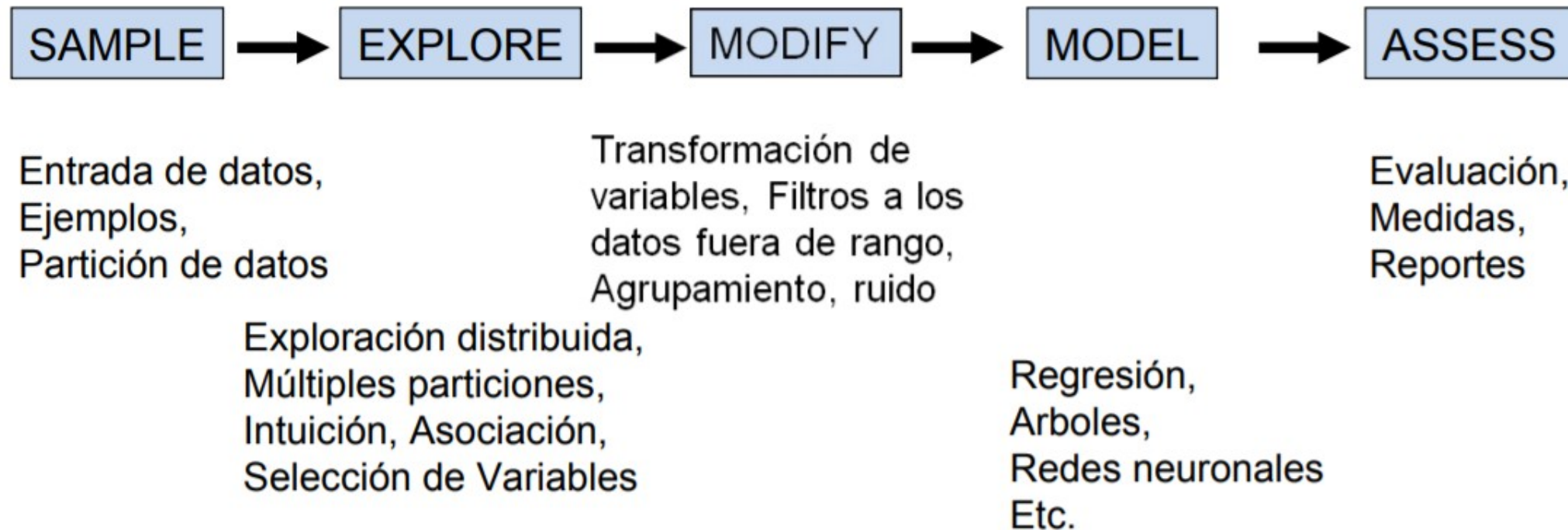
KDD



SEMMA

- SEMMA es el acrónimo a las cinco fases: (Sample, Explore, Modify, Model, Assess) La metodología es propuesta por SAS Institute Inc, la define como: “... proceso de selección, exploración y modelamiento de grandes cantidades de datos para descubrir patrones de negocios desconocidos...”

SEMMA



CRISP- DM

- Cross-Industry Standard Process for Data Mining (CRISPDM)
- Iniciativa financiada por la Comunidad Europea ha unido para desarrollar una plataforma para Minería de Datos.
- Objetivos:
 - – Fomentar la interoperabilidad de las herramientas a través de todo el proceso de minería de datos
 - – Eliminar la experiencia misteriosa y costosa de las tareas simples de minería de datos.

CRISP- DM

- **1. Comprensión del negocio:**
 - Entendimiento de los objetivos y requerimientos del proyecto.
 - Definición del problema de Minería de Datos
- **2. Comprensión de los datos**
 - Obtención conjunto inicial de datos.
 - Exploración del conjunto de datos.
 - Identificar las características de calidad de los datos ? Identificar los resultados iniciales obvios.
- **3. Preparación de Datos**
 - Selección de datos
 - Limpieza de datos

CRISP- DM

- **4. Modelamiento**

- Implementación en herramientas de Minería de Datos

- **5. Evaluación**

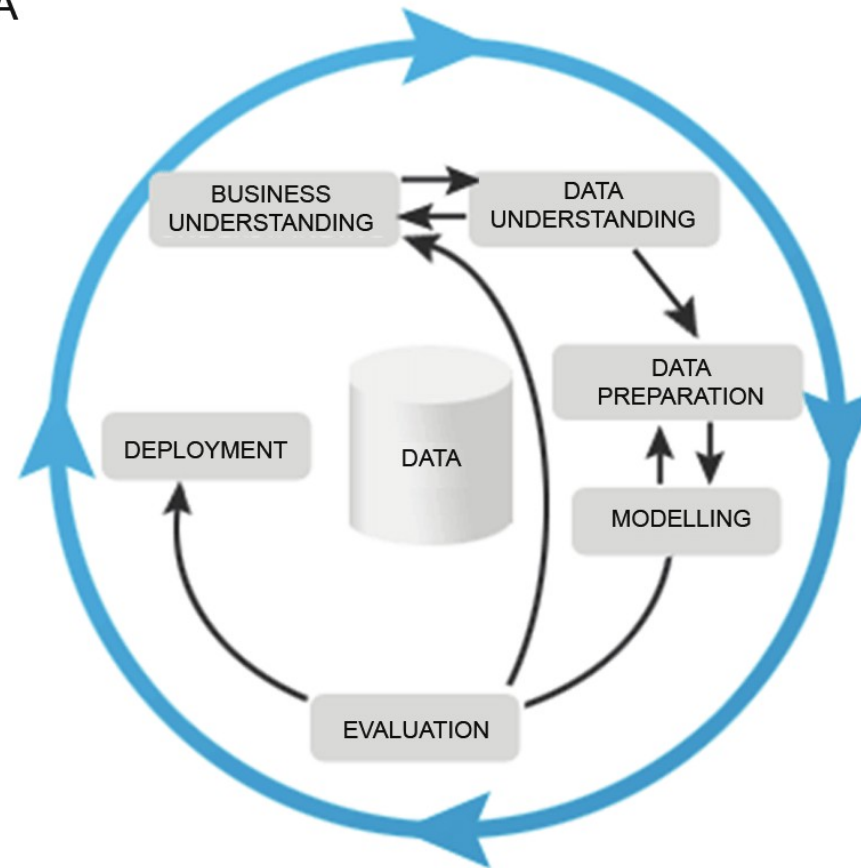
- Determinar si los resultados coinciden con los objetivos del negocio
- Identificar las temas de negocio que deberían haberse abordado

- **6. Despliegue**

- Instalar los modelos resultantes en la práctica
- Configuración para minería de datos de forma repetida ó continua

CRISP- DM

A



Comparación entre KDD, SEMMA y CRISP-DM

KDD	SEMMA	CRISP-DM
Pre KDD	xxxxx	Conocimiento del negocio
Selección	muestra	Conocimiento de los datos
Preprocesamiento	exploración	
Transformación	Modificación	
Minería de datos	Modelo	
interpretación / evaluación	evaluación	
Post KDD		

Casos de Estudio

Ejercicio (15 min)

Lee y discute los siguientes casos de éxito (Walmart, Bimbo, Amazon) y discute con en tu grupo cual de los procesos KDD, SEMMA y CRISP-DM mejor aplican a cada uno de los casos

(en otras palabras ¿cual usarian ustedes para ejecutar la asignación?)

Un caso de éxito de ERP: Walmart

- La clave del éxito de esta gran compañía se debe a su excelente operación logística por lo que ha sabido manejar muy bien sistemas estratégicos de información como el ERP. Walmart integró un sistema de información universal y lo expandió a todos sus colaboradores estratégicos, proveedores e intermediarios para que conocieran en tiempo real cómo estaban rotando sus mercancías en los puntos de ventas de la cadena de cada país, de cada ciudad.
- De ésta manera el proveedor o los suministradores podían conocer y planificar el próximo aprovisionamiento sin causar desabastos en el supermercado e inconvenientes con el cliente.

Un caso de éxito de ERP: Walmart

- Pero para lograr esta sincronización efectiva en el suministro, Walmart disponía dentro de sus políticas internas, pactar dentro de los contratos con sus proveedores, el compromiso de que éstos mismos por obligación tenían que integrar en sus operaciones, sistemas de ERP para que se pueda hacer las lecturas en tiempo real de las necesidades que se generaban en el punto.
- Actualmente, Walmart mantiene su estrategia adaptándola cada día más a las exigencias de los consumidores y que a pesar de compartir información crucial de sus operaciones con sus aliados estratégicos, queda por sentando que no siempre los secretos son las claves del éxito

Un caso de éxito de SCM: Bimbo

- Esta compañía creó una nueva área funcional de SCM con el objetivo de unificar y optimizar la cadena de suministro y garantizar la máxima frescura de los productos, el mejor servicio al cliente y la rentabilidad de la gestión logística e industrial.
- La empresa necesitaba una herramienta para planificar la cadena de suministro de los productos de vida media y larga (pastelería, tostados y aperitivos y golosinas).

Un caso de éxito de SCM: Bimbo

- Para ello, el área de SCM procedió a realizar un estudio funcional y organizativo con el objetivo de optimizar la cadena de suministro, conseguir a la vez un frescor óptimo de los productos, disminuir los stocks intermedios en la cadena, optimizar la producción y mejorar el servicio a las delegaciones de venta desde las fábricas.
- Se puede decir que Bimbo integra una adecuada implementación de software específico (SAP), como plataforma para la herramienta SCM, sincronizando y planificando sus componentes básicos: planeación de la demanda (DP), planificación de la producción (SNP), y la gestión de aprovisionamiento logístico.

Un caso de éxito de CRM: Amazon

- Jeff Bezos fue su fundador en 1994. Comenzó como una librería en línea que ofrecía libros en Estados Unidos y otros 45 países. Gracias a sus negocios virtuales pudo ofrecer 4 veces más de lo que una librería tradicional ofrecía por esos tiempos.
- Según la American Consumer Association, Amazon tuvo un 88% de clientes satisfechos con el servicio proporcionado, en los años 2002 y 2003, una cifra altísima para una empresa de servicios que no trata cara a cara con los cliente

Un caso de éxito de CRM: Amazon

- Amazon contempla su éxito estratégico en la adopción de los medios digitales, puesto ve en ellos un canal más efectivo para llegar a muchos más clientes de forma más rápida.
- Así mismo establece que por medio del mundo digital se podría ofrecer muchos más productos que en forma física, y es un hecho, puesto que en el mundo virtual las fronteras de las bodegas de almacenaje no existen metafóricamente hablando.

Bibliografía

- From Data Mining to Knowledge Discovery in Databases, Usama Fayyad, Gregory Piatetsky-Shapiro, From Data Mining to Knowledge Discovery in Databases and Padhraic Smyth, American Association for Artificial Intelligence, 1996 2.
- <http://www.crisp-dm.org/CRISPWP-0800.pdf> 3.
- <http://www.sas.com/offices/europe/uk/technologies/analytics/datamining/miner/semma.html>
- Anderson, C. (2008). The Long Tail, Whay the future of business Is Selling less of More.
- Arana, A. (2001). El impacto de internet en la transformacion de las relaciones con los clientes (CRM). Harvard Deusto. Marketing & Ventas, 70.
- Arjonilla, D., & Medina, G, J. A. (2013). La gestión de los sistemas de información en la empresa. Teoría y casos prácticos. Ediciones Pirámide.
- Ayuso, S., & Rodríguez, A, V. M. (2011). manual soluciones CRM. Formación empleo. Editorial CEP.
- Berenguer, J. M., & Ramos-Izquierdo, J. A. (2003). Negocios digitales: competir usando tecnologías de información.