



INICIO GRABACIÓN

MINERIA DE DATOS//DATAWAREHOUSE
IMPLEMENTACION MINERIA DE DATOS-
DATAWAREHOUSE



SANJOSÉ
FUNDACIÓN DE EDUCACIÓN SUPERIOR



INDICE

1

**IMPLEMENTACION MINERIA DE
DATOS-DATAWAREHOUSE**

2

CONCEPTOS

3

EJERCICIOS

4

CONCLUSIONES



SANJOSÉ
FUNDACIÓN DE EDUCACIÓN SUPERIOR

PRESENTACIÓN



KDD y Minería de Datos

Hoy en día, la cantidad de datos que ha sido almacenada en las bases de datos excede nuestra habilidad para reducir y analizar los datos sin el uso de técnicas de análisis automatizadas. Muchas bases de datos comerciales transaccionales y científicas crecen a una proporción fenomenal.

KDD [Knowledge Discovery in Databases] es 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. KDD se ha definido como “el proceso no trivial de identificación en los datos de patrones válidos, nuevos, potencialmente útiles, y finalmente comprensibles” . Se trata de interpretar grandes cantidades de datos y encontrar relaciones o patrones. Para conseguirlo harán falta técnicas de aprendizaje automático [Machine Learning] , estadística , bases de datos , técnicas de representación del conocimiento, razonamiento basado en casos [CBR, Case Based Reasoning], razonamiento aproximado, adquisición de conocimiento, redes de neuronas y visualización de datos. Tareas comunes en KDD son la inducción de reglas, los problemas de clasificación y clustering, el reconocimiento de patrones, el modelado predictivo, la detección de dependencias.

KDD es un campo creciente: hay muchas metodologías del descubrimiento del conocimiento en uso y bajo desarrollo. Algunas de estas técnicas son genéricas, mientras otros son de dominio específico.



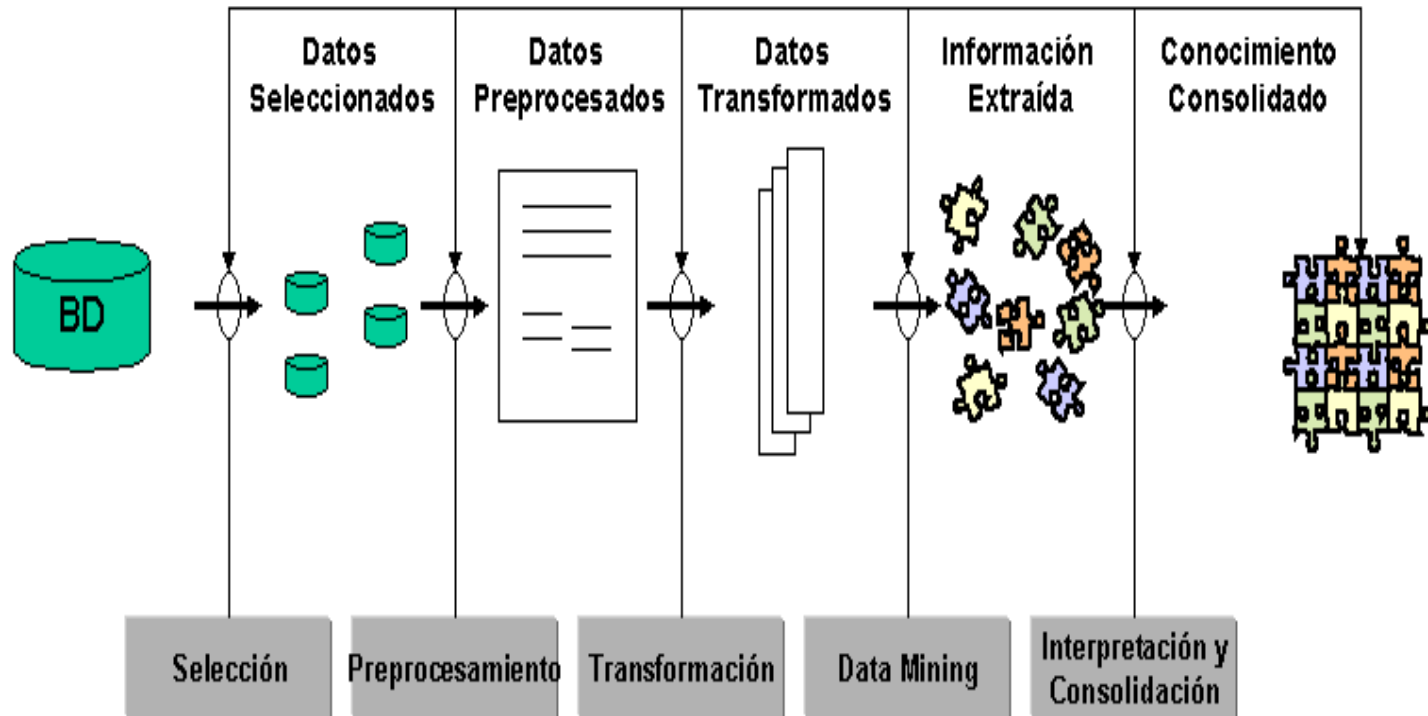
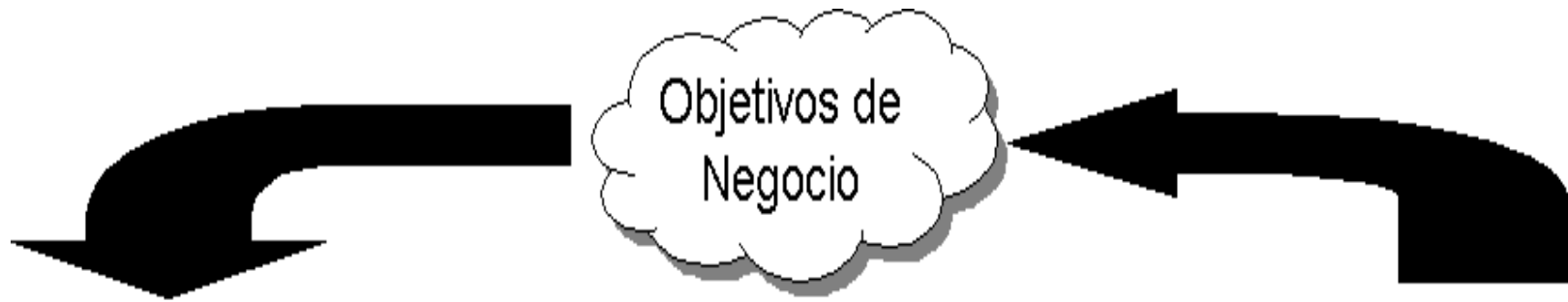
Los datos recogen un conjunto de hechos (una base de datos) y los patrones son expresiones que describen un subconjunto de los datos (un modelo aplicable a ese subconjunto). KDD involucra un proceso iterativo e interactivo de búsqueda de modelos, patrones o parámetros. Los patrones descubiertos han de ser válidos, novedosos para el sistema (para el usuario siempre que sea posible) y potencialmente útiles.

Se han de definir medidas cuantitativas para los patrones obtenidos (precisión, utilidad, beneficio obtenido...). Se debe establecer alguna medida de interés [interestingness] que considere la validez, utilidad 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.

El proceso de KDD

El proceso de KDD se inicia con la identificación de los datos. Para ello hay que imaginar qué datos se necesitan, dónde se pueden encontrar y cómo conseguirlos. Una vez que se dispone de datos, se deben seleccionar aquellos que sean útiles para los objetivos propuestos. Se preparan, poniéndolos en un formato adecuado.

Una vez se tienen los datos adecuados se procede a la minería de datos, proceso en el que se seleccionarán las herramientas y técnicas adecuadas para lograr los objetivos pretendidos. Y tras este proceso llega el análisis de resultados, con lo que se obtiene el conocimiento pretendido.





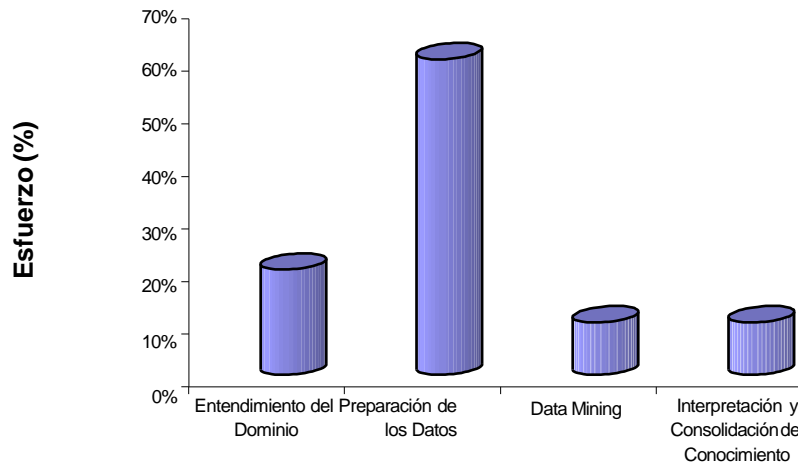
KDD es un proceso interactivo e iterativo, que involucra numerosos pasos e incluye muchas decisiones que deben ser tomadas por el usuario, y se estructura en las siguientes etapas:

- Comprensión del dominio de la aplicación, del conocimiento relevante y de los objetivos del usuario final.
- Creación del conjunto de datos: consiste en la selección del conjunto de datos, o del subconjunto de variables o muestra de datos, sobre los cuales se va a realizar el descubrimiento.
- Limpieza y preprocesamiento de los datos: Se compone de las operaciones, tales como: recolección de la información necesaria sobre la cual se va a realizar el proceso, decidir las estrategias sobre la forma en que se van a manejar los campos de los datos no disponibles, estimación del tiempo de la información y sus posibles cambios.
- Reducción de los datos y proyección: Encontrar las características más significativas para representar los datos, dependiendo del objetivo del proceso. En este paso se pueden utilizar métodos de transformación para reducir el número efectivo de variables a ser consideradas o para encontrar otras representaciones de los datos.
- Elegir la tarea de Minería de Datos: Decidir si el objetivo del proceso de KDD es: Regresión, Clasificación, Agrupamiento, etc.
- Elección del algoritmo(s) de Minería de Datos: Selección del método(s) a ser utilizado para buscar los patrones en los datos. Incluye además la decisión sobre que modelos y parámetros pueden ser los más apropiados.



Minería de Datos: Consiste en la búsqueda de los patrones de interés en una determinada forma de representación o sobre un conjunto de representaciones, utilizando para ello métodos de clasificación, reglas o árboles, regresión, agrupación, etc.

- Interpretación de los patrones encontrados. Dependiendo de los resultados, a veces se hace necesario regresar a uno de los pasos anteriores.
- Consolidación del conocimiento descubierto: consiste en la incorporación de este conocimiento al funcionamiento del sistema, o simplemente documentación e información a las partes interesadas.





Minería de Datos

Minería de Datos es un término genérico que engloba resultados de investigación, técnicas y herramientas usadas para extraer información útil de grandes bases de datos. Si bien Minería de Datos es una parte del proceso completo de KDD, en buena parte de la literatura los términos Minería de Datos y KDD se identifican como si fueran lo mismo. Concretamente, el término Minería de Datos es usado comúnmente por los estadísticos, analistas de datos, y por la comunidad de administradores de sistemas informáticos como todo el proceso del descubrimiento, mientras que el término KDD es utilizado más por los especialistas en Inteligencia Artificial.

El análisis de la información recopilada (por ejemplo, en un experimento científico) es habitual que sea un proceso completamente manual (basado por lo general en técnicas estadísticas). Sin embargo, cuando la cantidad de datos de los que disponemos aumenta la resolución manual del problema se hace intratable. Aquí es donde entra en juego el conjunto de técnicas de análisis automático al que nos referimos al hablar de Minería de Datos o KDD.

Hasta ahora, los mayores éxitos en Minería de Datos se pueden atribuir directa o indirectamente a avances en bases de datos (un campo en el que los ordenadores superan a los humanos). No obstante, muchos problemas de representación del conocimiento y de reducción de la complejidad de la búsqueda necesaria (usando conocimiento a priori) están aún por resolver. Ahí reside el interés que ha despertado el tema entre investigadores de todo el mundo.



Tecnologías de Apoyo

Razonamiento estadístico

Las técnicas y métodos estadísticas del razonamiento han sido utilizados durante varias décadas, siendo los únicos medios de analizar los datos en el pasado. Numerosos paquetes están ahora disponibles para computar promedios, sumas, y diferentes distribuciones para diferentes aplicaciones. Por ejemplo, la oficina del censo usa análisis y métodos estadísticos para analizar la población en un país. Más recientemente, las técnicas estadísticas del razonamiento están jugando un papel importante en la Minería de Datos. Algunos paquetes estadísticos que han sido utilizados durante mucho tiempo, se han integrado con las diferentes bases de datos, y se están comercializándose en la actualidad como productos para la Minería de Datos.

Visualización

Las tecnologías de la visualización muestran gráficamente los datos en las bases de datos. Se ha investigado mucho sobre la visualización y el campo ha adelantado un gran trecho sobre todo con la incorporación de la informática multimedia. Por ejemplo, los datos en las bases de datos serán filas y filas de valores numéricos, y las herramientas de visualización toman estos datos y trazan con ellos algún tipo de gráfico. 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, y algunos trabajos sobre este tema están recogidos



Procesamiento paralelo

El procesamiento paralelo es una técnica que ha sido utilizado durante mucho tiempo. El área se ha desarrollado significativamente, desde sistemas con un único procesador hasta sistemas multiprocesador. 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. Hay muchos trabajos sobre la utilización de las arquitecturas paralelas para el procesamiento de las bases de datos (véase, por ejemplo. A pesar de haberse realizado considerable trabajo sobre el tema, estos sistemas no fueron comercializados hasta el desarrollo del *data warehouse*, ya que muchos de los *data warehouses* emplean el procesamiento paralelo para acelerar el proceso de las consultas.

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, y se basan en la teoría de la decisión. Se puede considerar a las herramientas de Minería de Datos como tipos especiales de herramientas de apoyo a la toma de decisiones. Las herramientas de apoyo a la toma de decisiones pertenecen a una amplia categoría.



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 la Minería de Datos.

El aprendizaje automático aparece continuamente en la realización de aprendizaje computacional desde la experiencia. Como Mitchell describe en su excelente texto sobre aprendizaje automático, el aprendizaje automático consiste en aprender de las experiencias del pasado con respecto a alguna medida de rendimiento. Por ejemplo, en las aplicaciones de los juegos de computadora, el aprendizaje automático podría ser aprender a jugar un juego de ajedrez, desde las experiencias del pasado que podrían ser juegos que el ordenador juega contra sí mismo, con respecto a alguna medida de rendimiento, como ganar un cierto número de partidas.



Áreas de Aplicación

En este punto se presentan las principales áreas y sectores empresariales en las que se puede aplicar la minería de datos.

Marketing

Actualmente con la generación de los puntos de ventas informatizados y conectados a un ordenador central, y el constante uso de las tarjetas de créditos se genera gran cantidad de información que hay que analizar. Con ello se puede emplear la minería de datos para:

- Identificar patrones de compra de los clientes: Determinar cómo compran, a partir de sus principales características, conocer el grado de interés sobre tipos de productos, si compran determinados productos en determinados momentos,...
- Segmentación de clientes: Consiste en la agrupación de los clientes con características similares, por ejemplo demográficas. Es una importante herramienta en la estrategia de marketing que permite realizar ofertas acordes a diferentes tipos de comportamiento de los consumidores.
- Predecir respuestas a campañas de *mailing*: Estas campañas son caras y pueden llegar a ser molestas para los clientes a los que no le interesan el tipo de producto promocionado por lo que es importante limitarlas a los individuos con una alta probabilidad de interesarse por el producto. Está por ello muy relacionada con la segmentación de clientes.
- Análisis de cestas de la compra [market-basket analysis]: Consiste en descubrir relaciones entre productos, esto es, determinar qué productos suelen comprarse junto con otros, con el fin de distribuirlos adecuadamente.



Compañías de Seguro

En el sector de las compañías de seguros y la salud privada, se pueden emplear las técnicas de minería de datos, por ejemplo para:

- Análisis de procedimientos médicos solicitados conjuntamente.
- Predecir qué clientes compran nuevas pólizas.
- Identificar patrones de comportamiento para clientes con riesgo.
- Identificar comportamiento fraudulento.

Banca

En el sector bancario la información que puede almacenarse es, además de las cuentas de los clientes, la relativa a la utilización de las tarjetas de crédito, que puede permitir conocer hábitos y patrones de comportamiento de los usuarios. Esta información puede aplicarse para:

- Detectar patrones de uso fraudulento de tarjetas de crédito.

Identificar clientes leales: Es importante para las compañías de cualquier sector mantener los clientes. Y es que hay estudios que demuestran que es cuatro veces más caros obtener nuevos clientes que mantener los existentes.

- Predecir clientes con probabilidad de cambiar su afiliación.
- Determinar gasto en tarjeta de crédito por grupos.
- Encontrar correlaciones entre indicadores financieros.
- Identificar reglas de mercado de valores a partir de históricos:

Telecomunicaciones



En el sector de las telecomunicaciones se puede almacenar información interesante sobre las llamadas realizadas, tal como el destino, la duración, la fecha,... en que se realiza la llamada, por ejemplo para:

- Detección de fraude telefónico: Mediante por ejemplo el agrupamiento o *clustering* se pueden detectar patrones en los datos que permitan detectar fraudes.

Medicina

También en el campo médico se almacena gran cantidad de información, sobre los pacientes, tal como enfermedades pasadas, tratamientos impuestos, pruebas realizadas, evolución,...

Se pueden emplear técnicas de minería de datos con esta información, por ejemplo, para:

- Identificación de terapias médicas satisfactorias para diferentes enfermedades.
- Asociación de síntomas y clasificación diferencial de patologías.
- Estudio de factores (genéticos, precedentes, hábitos, alimenticios,...) de riesgo para la salud en distintas patologías.
- Segmentación de pacientes para una atención más inteligente según su grupo.
- Estudios epidemiológicos, análisis de rendimientos de campañas de información, prevención, sustitución de fármacos,...
- Identificación de terapias médicas y tratamientos erróneos para determinadas enfermedades.

Industria farmacéutica



En el sector químico y farmacéutico se almacenan gran cantidad de información:

- Bases de datos de dominio público conteniendo información sobre estructuras y propiedades de componentes químicos.
- Resultados de universidades y laboratorios publicadas en revistas técnicas.
- Datos generados en la realización de los experimentos.
- Datos propios de la empresa.

Los datos son almacenados en diferentes categorías y a cada categoría se le aplica un diferente trato. Se podrían realizar, entre otras, las siguientes operaciones con la información obtenida:

- *Clustering* de moléculas: Consiste en el agrupamiento de moléculas que presentan un cierto nivel de similitud, con lo que se pueden descubrir importantes propiedades químicas.
- Búsqueda de todas las moléculas que contienen un patrón específico: Se podría introducir una subestructura (un patrón), devolviendo el sistema todas las moléculas que son similares a dicha estructura.
- Búsqueda de todas las moléculas que vincula un camino específico hacia una molécula objetivo: Realizar una búsqueda exhaustiva puede ser impracticable, por lo que se pueden usar restricciones en el espacio de búsqueda.
- Predicción de resultado de experimentos de una nueva molécula a partir de los datos almacenados: A través de determinadas técnicas de inteligencia artificial es posible predecir los resultados a nuevos experimentos a partir de los datos, con el consiguiente ahorro de tiempo y dinero.



Biología

Con la finalización en los próximos años del Proyecto Genoma Humano y el almacenamiento de toda la información que está generando en bases de datos accesibles por Internet, el siguiente reto consiste en descubrir cómo funcionan nuestros genes y su influencia en la salud. Existen nuevas tecnologías (chips de ADN, proteómica, genómica funcional, variabilidad genética individual) que están posibilitando el desarrollo de una “nueva biología” que permite extraer conocimiento biomédicos a partir de bases de datos experimentales en el entorno de un ordenador básicamente mediante técnicas de minería de datos y visualización. Estos trabajos forman parte de los desarrollos de la Bioinformática.

Tendencias de la Minería de Datos

El interés que despierta la Minería de Datos para el análisis de la información especialmente en el área comercial hace que se busquen nuevas aplicaciones basadas en esta tecnología. Algunas de las principales nuevas aplicaciones basadas en la Minería de Datos se presentan a continuación.

Minería de Textos

La Minería de Textos [Text Mining] surge ante el problema cada vez más apremiante de extraer información automáticamente a partir de masas de textos. Se trata así de extraer información de datos no estructurados: texto plano.



Existen varias aproximaciones a la representación de la información no estructurada:

- “Bag of Words”: Cada palabra constituye una posición de un vector y el valor corresponde con el número de veces que ha aparecido.
- *N*-gramas o frases: Permite tener en cuenta el orden de las palabras. Trata mejor frases negativas “... *excepto* ...”, “... *pero no* ...”, que tomarían en otro caso las palabras que le siguen como relevantes.
- Representación relacional (primer orden): Permite detectar patrones más complejos (si la palabra *X* está a la izquierda de la palabra *Y* en la misma frase...).
- Categorías de conceptos.

Casi todos se enfrentan con el “vocabulary problem”: Tienen problemas con la sinonimia, la polisemia, los lemas, etc.

Un ejemplo de aplicación basada en Minería de Textos es la generación automática de índices en documentos. Otras más complicadas consistirían en escanear completamente un texto y mostrar un mapa en el que las partes más relacionadas, o los documentos más relacionados se coloquen cerca unos de otros. En este caso se trataría de analizar las palabras en el contexto en que se encuentren.

En cualquier caso, aunque aún no se ha avanzado mucho en el área de Minería de Textos, ya hay productos comerciales que emplean esta tecnología con diferentes propósitos.

Minería de datos Web

La Minería de datos Web [Web Mining] es una tecnología usada para descubrir conocimiento interesante en todos los aspectos relacionados a la Web. Es uno de los mayores retos. El enorme volumen de datos en la Web generado por la explosión de usuarios y el desarrollo de librerías digitales hace que la extracción de la información útil sea un gran problema



Minería de Datos y Almacenamiento de Datos

Como se ha enfatizado repetidamente, los datos son críticos para hacer data mining. Por consiguiente, se necesitan sistemas de bases de datos para manejar los datos a los que aplicar data mining eficazmente. Estos sistemas podrían ser sistemas de data warehouse o sistemas de bases de datos.

Arquitectura, Modelado, Diseño, y Aspectos de la Administración

Las técnicas de data mining existen desde hace algún tiempo. ¿Por qué entonces data mining se ha hecho tan popular ahora? La principal razón es que ahora con los sistemas de bases de datos se pueden representar, almacenar y recuperar los datos, y reforzar características como la integridad y seguridad.

Data mining y Funciones de Bases de datos

En el caso de integración fuerte entre el SGBD y data mining hay un fuerte impacto sobre las diferentes funciones del sistema de bases de datos. Por ejemplo, en el procesamiento de consultas. Se han realizado trabajos para examinar lenguajes de consultas como SQL y determinar si se necesitan extensiones para soportar data mining. Si hay estructuras adicionales y consultas que son complejas, entonces el optimizador de consultas tiene que ser adaptado para manejar esos casos. Estrechamente relacionado con la optimización de consultas esta la eficiencia de las estructuras de almacenamiento, índices, y métodos de acceso. Pueden ser necesarios mecanismos especiales para apoyar data mining en el procesamiento de consultas.



DATA WAREHOUSE

Por definición, las cargas de trabajo del data warehouse están destinadas para el apoyo a la toma de decisiones y por lo tanto, tienen consultas intensivas (con actividades ocasionales de inserción por lotes); asimismo, los propios data warehouses tienden a ser bastante grandes (a menudo mayores que 500GB y con una tasa de crecimiento de hasta el 50 por ciento anual). Por consecuencia, es difícil -aunque no imposible- perfeccionar el rendimiento. También puede ser un problema la escalabilidad. Contribuyen a ese problema

(a) los errores de diseño de la base de datos, (b) el uso ineficiente de los operadores relacionales, (c) la debilidad en la implementación del modelo relacional del DBMS, (d) la falta de escalabilidad del propio **DBMS** y (e) los errores de diseño arquitectónico que limitan la capacidad e imposibilitan la escalabilidad de la plataforma.

• APLICACIONES DE LOS DATA WAREHOUSE

La explotación del Data Warehouse puede realizarse mediante diversas técnicas:

- Query & Reporting
- On-line analytical processing (OLAP)
- Executive Information System (EIS)
- Decision Support Systems (DSS)
- Visualización de la información
- Data Mining ó Minería de Datos



DATA WAREHOUSE Y DATA MINING

Data warehouse almacena los datos de las bases de datos heterogéneas para que los usuarios consulten sólo un único aspecto. Las respuestas que un usuario consigue a una consulta dependen de los volúmenes del data warehouse. El data warehouse en general no intenta extraer la información de los datos almacenados. Data warehouse estructura y organiza los datos para soportar funciones de administración, data mining intenta extraer la información útil, así como predecir las tendencias de los datos. La relación entre el data warehouse y data mining. Observe que no es necesario construir un data warehouse para hacer data mining, ya que también puede aplicarse data mining a las bases de datos. Sin embargo, un data warehouse estructura los datos de tal manera que facilita data mining, por lo que en muchos casos es muy deseable tener un almacén de los datos para llevar a cabo data mining..

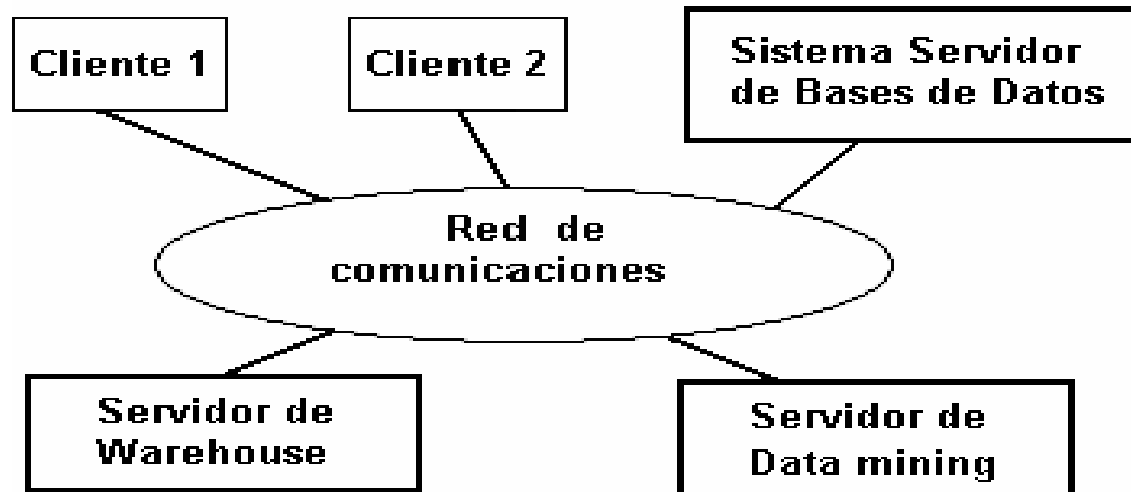


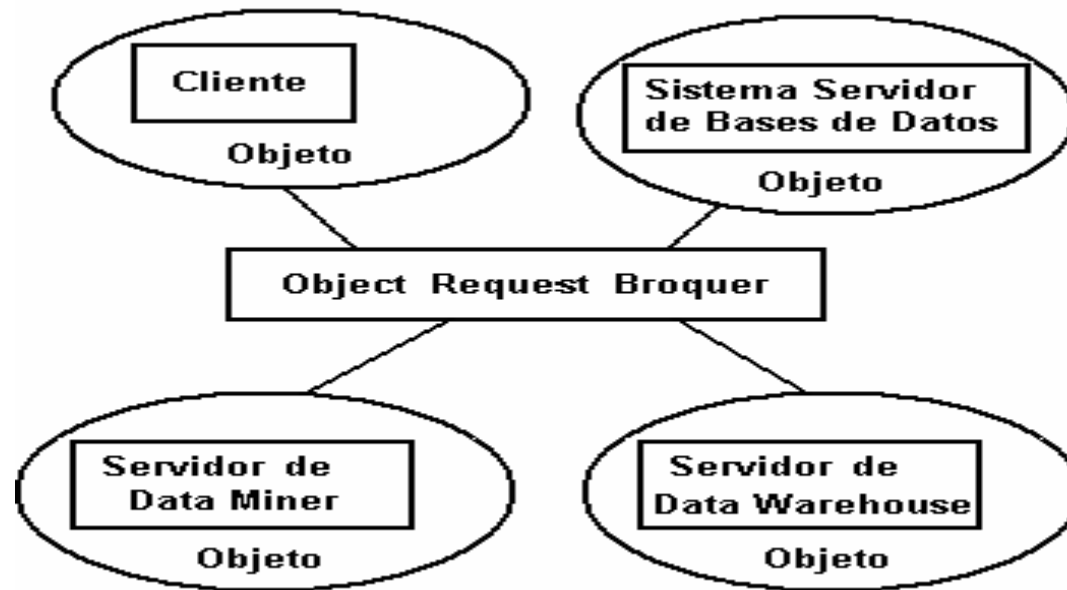
Esencialmente, un warehouse organiza los datos eficazmente para realizar data mining sobre ellos. La pregunta es entonces ¿Es imprescindible tener un warehouse para hacer data mining? La respuesta es que es muy interesante tener un warehouse, pero esto no significa que sea imprescindible. Podría usarse un buen SGBD para gestionar una base de datos eficazmente. También, a menudo con un warehouse no se tienen datos transaccionales. Por lo tanto, los datos no pueden ser actuales, y los resultados obtenidos desde data mining tampoco lo serán. Si se necesita la información actualizada, entonces se podría hacer data mining sobre una base de datos administrada por un SGBD que también tenga características de procesamiento de transacciones. Hacer data mining sobre datos que se actualizan a menudo es un desafío. Típicamente data mining se ha usado sobre los datos de apoyo a la toma de decisiones. 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. De momento al menos, es crítico



El Data Mining en la Arquitectura del Sistema

Considere la arquitectura de la Figura . En este ejemplo, la herramienta de data mining podría usarse como un servidor, los sistemas de administración de bases de datos podrían ser otro servidor, mientras el data warehouse sería un tercer servidor. El cliente emite las peticiones al sistema de base de datos, al warehouse, y al componente de data mining

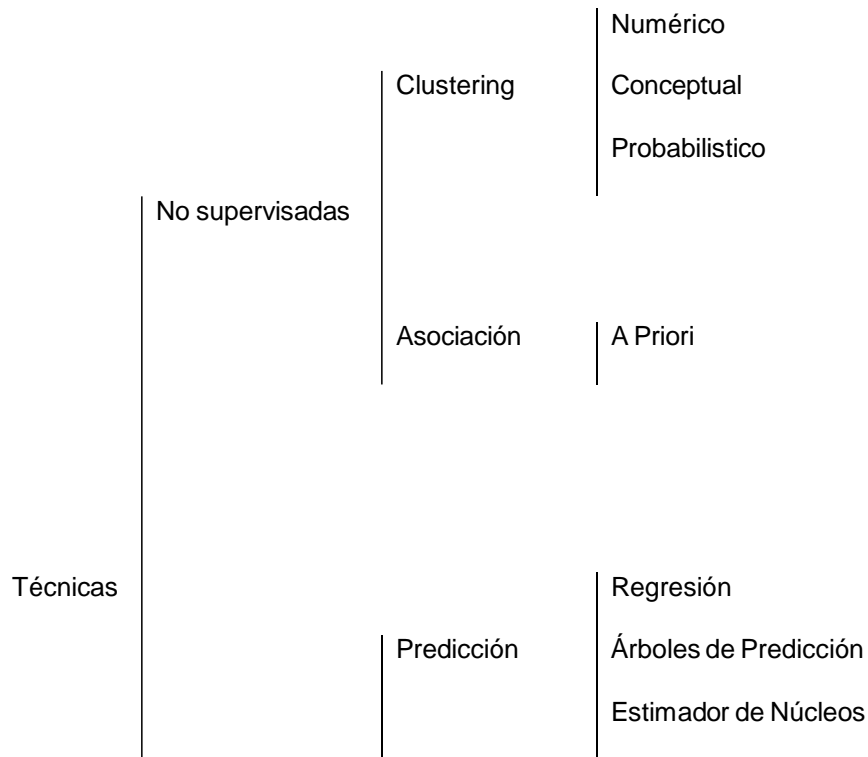






Técnicas de Minería de Datos

Como ya se ha comentado, las técnicas de Minería de Datos (una etapa dentro del proceso completo de KDD) intentan obtener patrones o modelos a partir de los datos recopilados. Decidir si los modelos obtenidos son útiles o no suele requerir una valoración subjetiva por parte del usuario. Las técnicas de Minería de Datos se clasifican en dos grandes categorías: supervisadas o predictivas y no supervisadas o descriptivas.





Supervisadas

Clasificación

Tabla de Decisión

Árboles de Decisión

Inducción de Reglas

Bayesiana

Basado en Ejemplares

Redes de Neuronas

Lógica Borrosa

Técnicas Genéticas



La predicción

Es el proceso que intenta determinar los valores de una o varias variables, a partir de un conjunto de datos. La predicción de valores continuos puede planificarse por las técnicas estadísticas de regresión. Por ejemplo, para predecir el sueldo de un graduado de la universidad con 10 años de experiencia de trabajo, o las ventas potenciales de un nuevo producto dado su precio. Se pueden resolver muchos problemas por medio de la regresión lineal, y puede conseguirse todavía más aplicando las transformaciones a las variables para que un problema no lineal pueda convertirse a uno lineal.

Regresión no lineal.

En muchas ocasiones los datos no muestran una dependencia lineal. Esto es lo que sucede si, por ejemplo, la variable respuesta depende de las variables independientes según una función polinómica, dando lugar a una regresión polinómica que puede planearse agregando las condiciones polinómicas al modelo lineal básico.

Árboles de Predicción

Los árboles de predicción numérica son similares a los árboles de decisión, que se estudiarán más adelante, excepto en que la clase a predecir es continua. En este caso, cada nodo hoja almacena un valor de clase consistente en la media de las instancias que se clasifican con esa hoja, en cuyo caso estamos hablando de un *árbol de regresión*, o bien un modelo lineal que predice el valor de la clase, en cuyo caso se habla de *árbol de modelos*. En el caso del algoritmo, se trata de obtener un árbol de modelos, si bien se puede utilizar para obtener un árbol de regresión, por ser éste un caso específico de árbol de modelos.

La clasificación



La clasificación es el proceso de dividir un conjunto de datos en grupos mutuamente excluyentes, de tal forma que cada miembro de un grupo esté lo mas cerca posible de otros y grupos diferentes estén lo más lejos posible de otros, donde la distancia se mide con respecto a las variables especificadas, que se quieren predecir.

Ejemplo	Vista	Temperatura	Humedad	Viento	Jugar
1	Soleado	Alta (85)	Alta (85)	No	No
2	Soleado	Alta (80)	Alta (90)	Sí	No
3	Nublado	Alta (83)	Alta (86)	No	Sí
4	Lluvioso	Media (70)	Alta (96)	No	Sí
5	Lluvioso	Baja (68)	Normal (80)	No	Sí
6	Lluvioso	Baja (65)	Normal (70)	Sí	No
7	Nublado	Baja (64)	Normal (65)	Sí	Sí
8	Soleado	Media (72)	Alta (95)	No	No
9	Soleado	Baja (69)	Normal (70)	No	Sí
10	Lluvioso	Media (75)	Normal (80)	No	Sí
11	Soleado	Media (75)	Normal (70)	Sí	Sí
12	Nublado	Media (72)	Alta (90)	Sí	Sí
13	Nublado	Alta (81)	Normal (75)	No	Sí
14	Lluvioso	Media (74)	Alta (91)	Sí	No

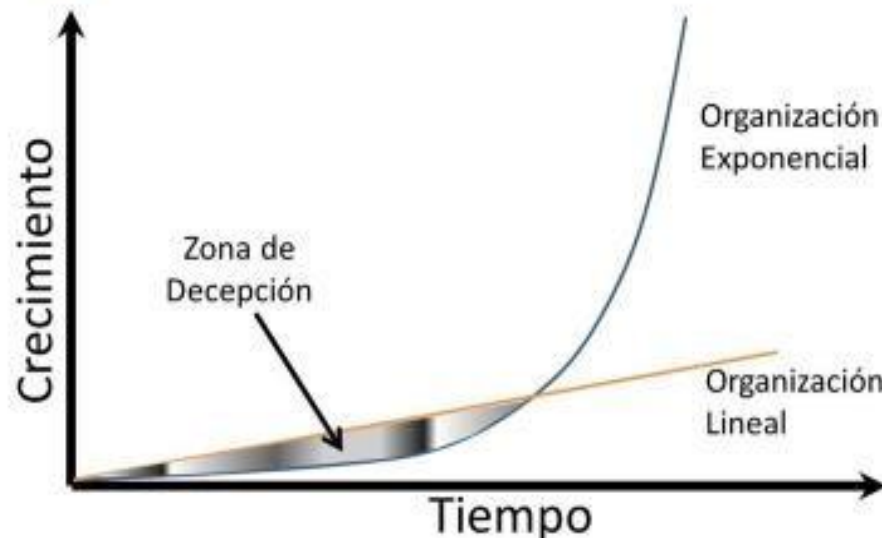
Bibliografía: An Introduction to Generalized Linear Models. New York : Chapman and Hall, 1990.

CASO

Organizaciones exponenciales



¿Qué son?



Una Organización cuyo impacto es desproporcionadamente grande, al menos 10 veces más que su competencia, debido al uso de nuevas técnicas organizativas que se enfocan en el uso de tecnologías de aceleración



Si crees que el ritmo de la innovación ha sido rápido en estos últimos años, prepárate para lo que viene porque aún no hemos visto nada”

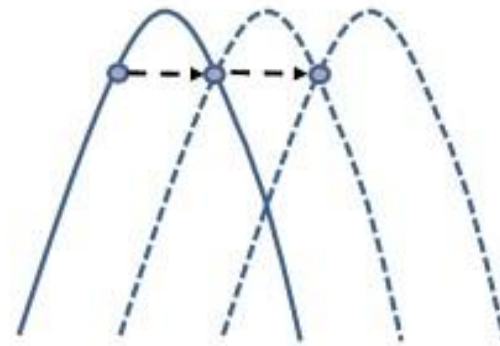
Peter H. Diamandis - cofundador y presidente ejecutivo de Singularity

<https://www.diamandis.com/>



Organización exponencial Transformación continua y “Prime time”

Los ciclos de vida de los modelos de negocio son cada vez más cortos.
Las empresas han de estar continuamente desarrollando nuevos modelos de negocio a través de la innovación, en un proceso de transformación continua,
¿Cómo pueden aprovechar las empresas las nuevas tecnologías digitales para esta finalidad?



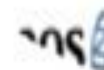


Organizaciones exponenciales

Una manifestación de doble capacidad:
disrupción y aceleración

Compañías con capacidad de generar
vinculación masiva de clientes en muy poco
tiempo, y paralelamente, presentar niveles
de crecimiento exponencial de su valor de
mercado

Cinco grandes de la Economía Digital: Apple,
Alphabet, Microsoft, Amazon y Facebook, se
han convertido en las cinco empresas más
valiosas del mundo.



ALICIA



Organizaciones exponenciales



- En vez de usar cientos de personas y enormes localizaciones físicas y edificios, se construyen sobre las tecnologías de la información.
- Se basan en 4 elementos claves:
 1. Las tecnologías de la información van a crecer más y más rápidamente (Ley de los rendimientos acelerados)
 2. El motor que impulsa todo este fenómeno es la información.
 3. El modelo se multiplica (dobla su valor),
 4. El proceso se hace imparable (crecimiento exponencial)
- Ya estamos en un mundo exponencial, sobre todo en algunos sectores, pero pronto, lo exponencial golpeará a todos los sectores desde diversos lados, aún cuando el foco del negocio sea tradicional.

© ALFONSO



Tecnologías digitales y empresas exponenciales



La aparición y desarrollo paralelo de varias tecnologías que, además de su capacidad disruptiva, están suponiendo la disponibilidad de capacidad casi ilimitada, a nivel de empresa y usuario individual, de obtención, procesamiento, almacenamiento, análisis, visualización, comunicación y compartición de información masiva útil a costes muy reducidos.

ALICIA



Digitalización



Permite una automatización y una rapidez que incrementa la escalabilidad exponencialmente.

EL MODELO 6Ds DE LAS ORGANIZACIONES EXPONENCIALES

"The Six Ds are a chain reaction of technological progression, a road map of rapid development that always leads to enormous upheaval and opportunity."

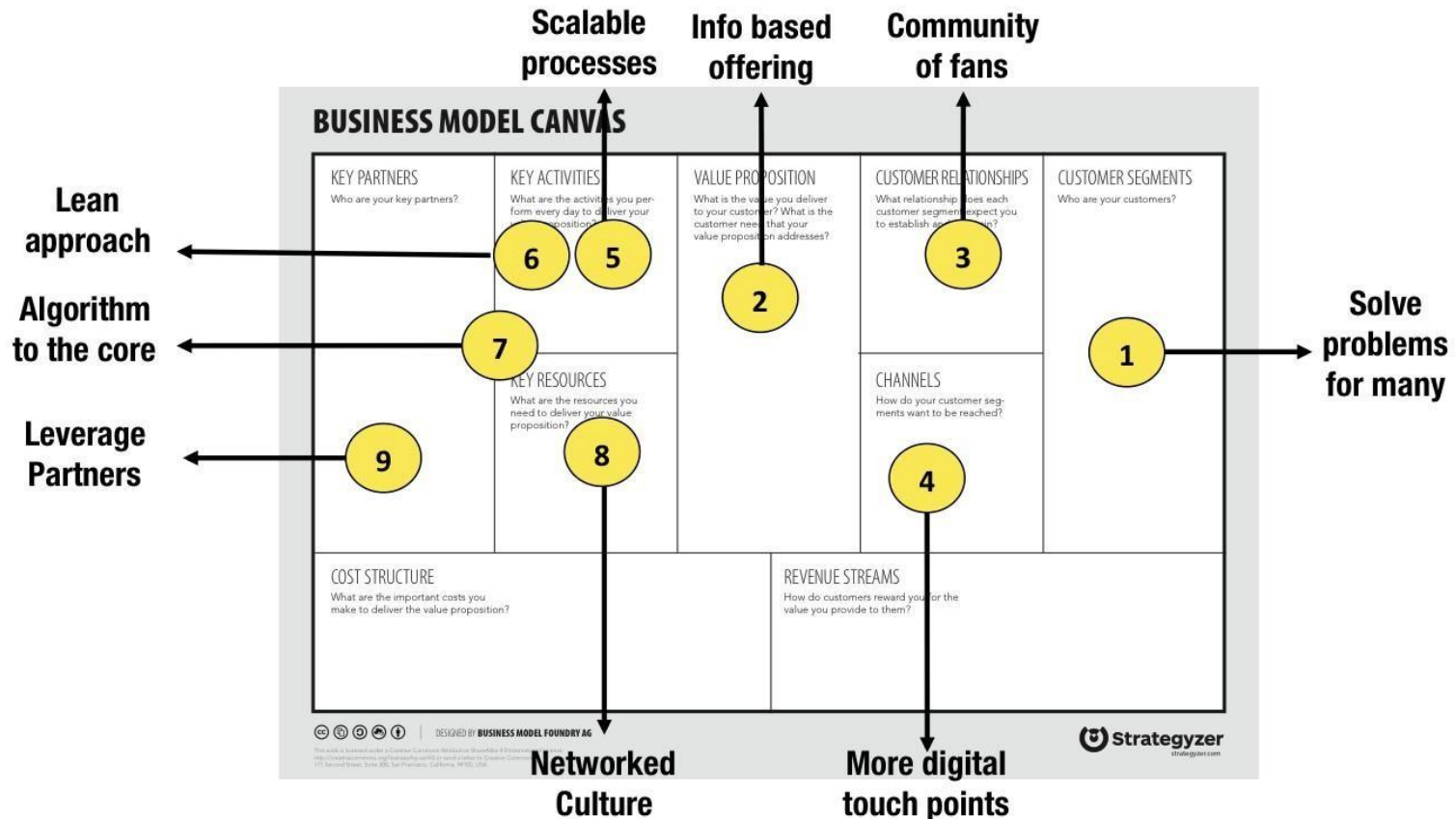
Peter Diamandis and Steven Kotler



Fuente: Steven Kotler. 6D's Model

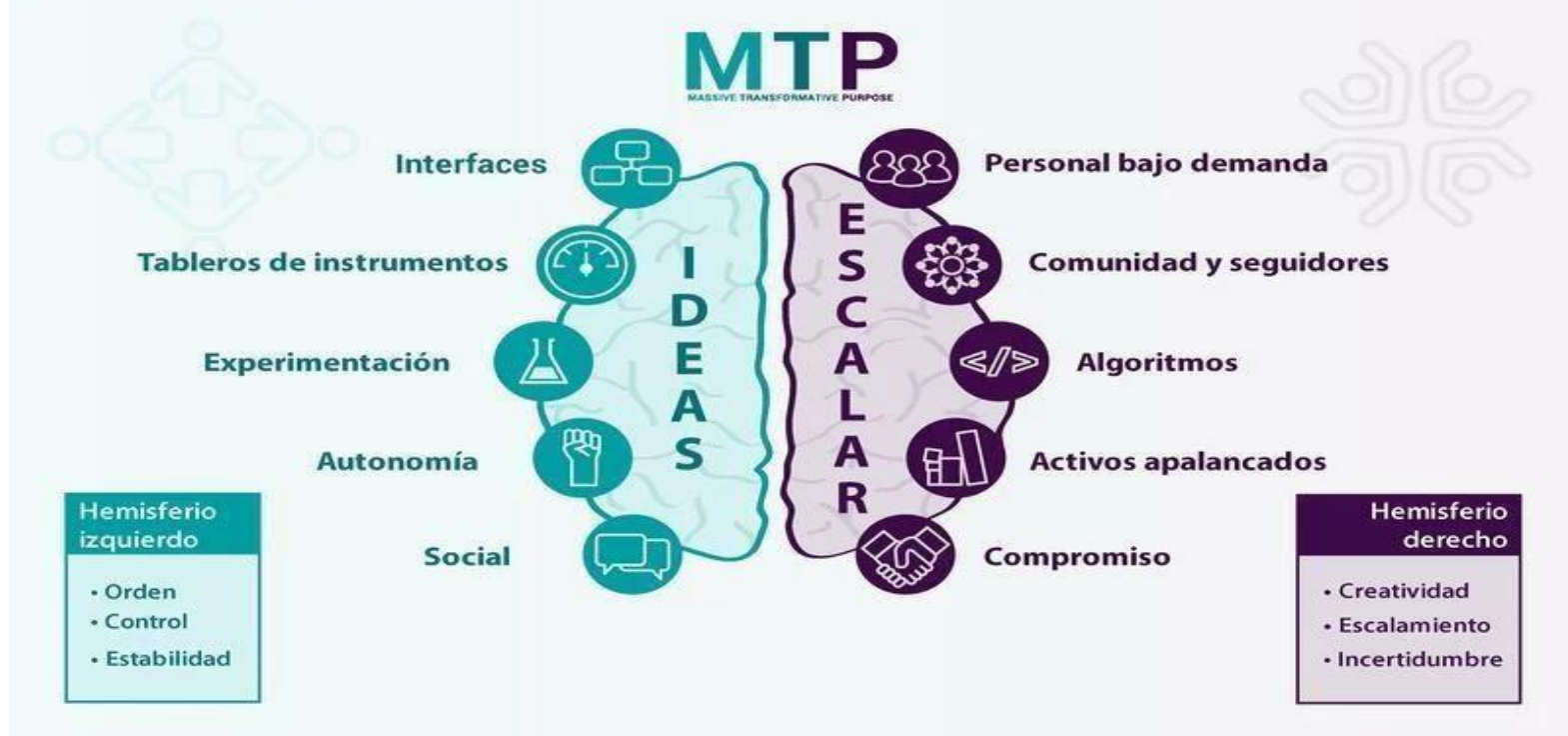
GuilleLorbada
@lorbada

Bibliografía: "An Introduction to Generalized Linear Models". New York: Chapman and Hall, 1990.



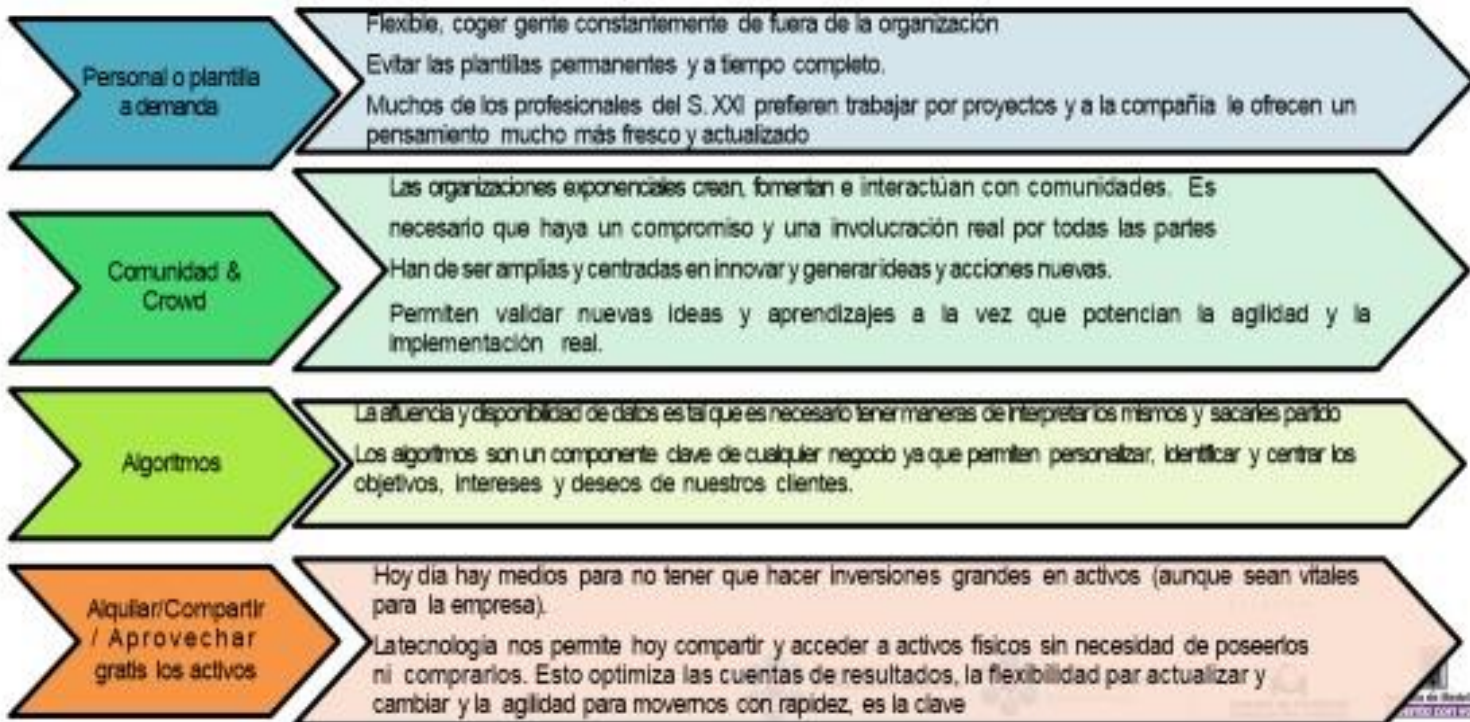
Bibliografía: An Introduction to Generalized Linear Models". New York: Chapman and Hall, 1990.

La fórmula ExO





Atributos externos (escalar)





Atributos internos (ideas)

PROCESOS QUE FILTRAN Y ENCAJAN LOS ELEMENTOS EXTERNOS CON LOS INTERNOS

- Cuanto más automatizados y menos manuales sean estos procesos, mejor, ya que es lo que permite a las empresas escalar con velocidad.

CUADRO DE MANDO

- Es necesario tener un cuadro de mando simple con todas las métricas necesarias. Hay que tener unos KPIs muy definidos que nos indiquen que estamos aportando valor.
- En las ExOs, los KPI se han sustituido por el método OKR (Objectives and Key Results)

EXPERIMENTACIÓN

- Implementación real de la metodología LEAN: Es necesaria una experimentación constante, un proceso de iteración para reducir el riesgo, una disposición a equivocarse y fallar y aprender de ello y tener una mentalidad de aprendizaje escalable constante.

AUTONOMÍA

- Es la siguiente fase del trabajo en equipo: Funcionar con equipos multidisciplinarios y una autoridad descentralizada, implica un cambio en el rol del manager.

TECNOLOGÍAS SOCIALES

- Crean interacciones horizontales incluso en las compañías más verticales. Se encargan de generar conexión, compromiso, confianza y transparencia.



Mejorar la eficiencia en todos los procesos internos



I+D, Producción, Comercialización,
Marketing, Talento Humano, Atención al
cliente o consumidor, Calidad, Finanzas.
Así como en la estructura y los sistemas de
gestión de la organización.

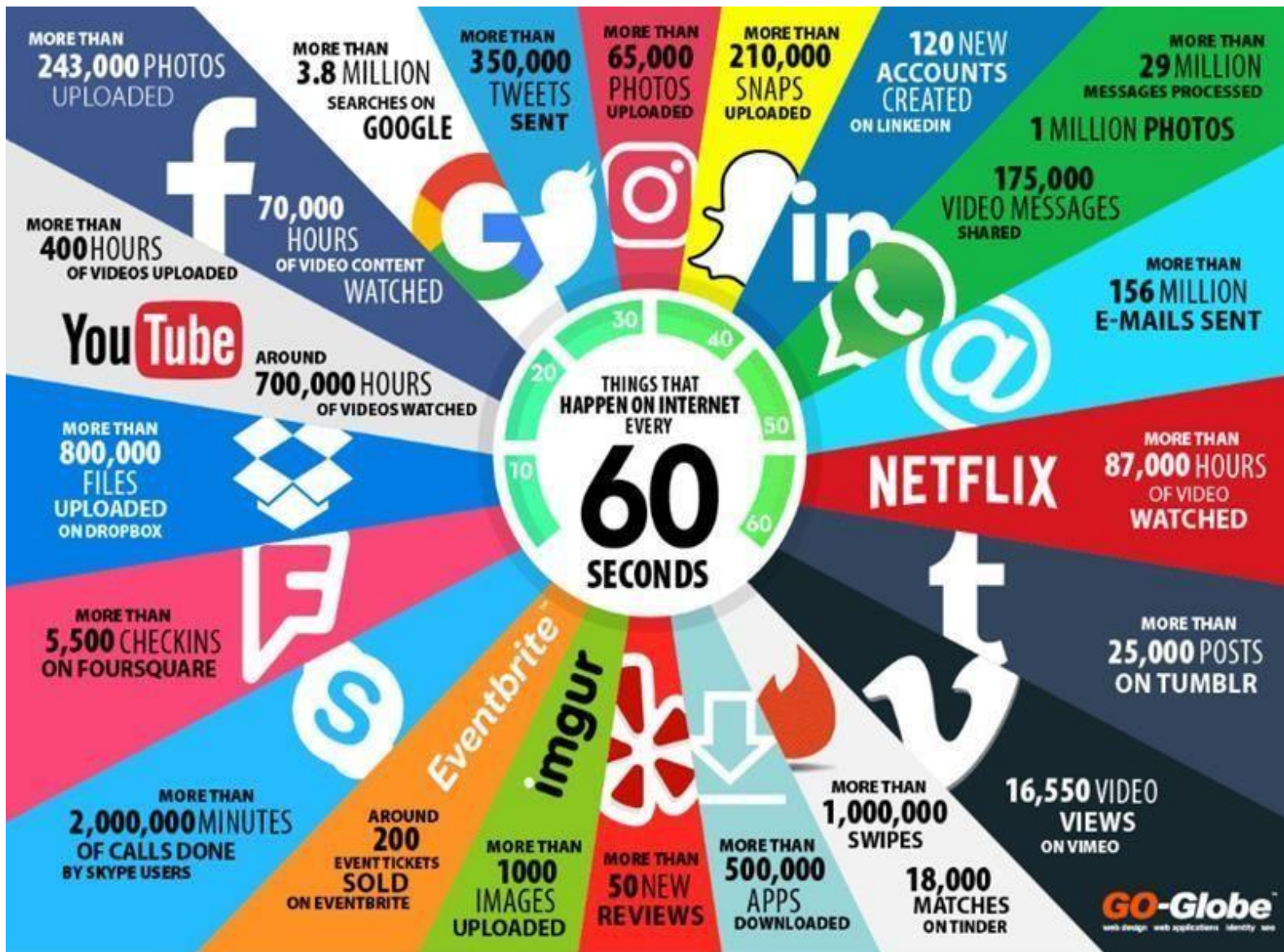


Profundizar en el análisis de



Impulsar el Big Data para convertir los datos en conocimiento y en decisión, trabajar y decidir tomando como base la información, mediante analítica predictiva.





Bibliografía: An Introduction to Generalized Linear Models". New York: Chapman and Hall, 1990.



Datos como activos y oportunidades



"Recolectar y organizar grandes cantidades de datos para hacer algo con ellos ... incluyendo la identificación de patrones, para contar historias que resalten ideas significativas."

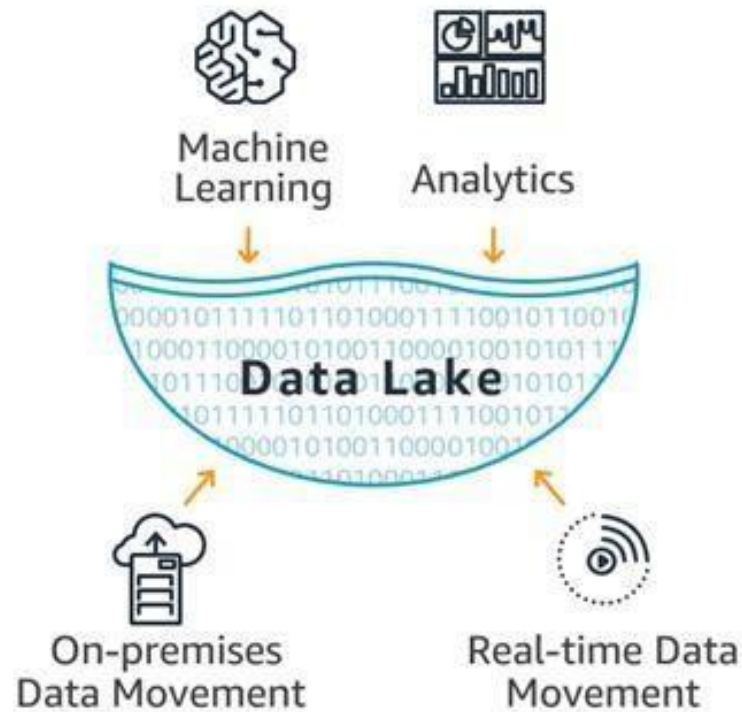


Data Vision

Accionar los datos para lograr una diferencia



Algunos conceptos básicos: Data lake



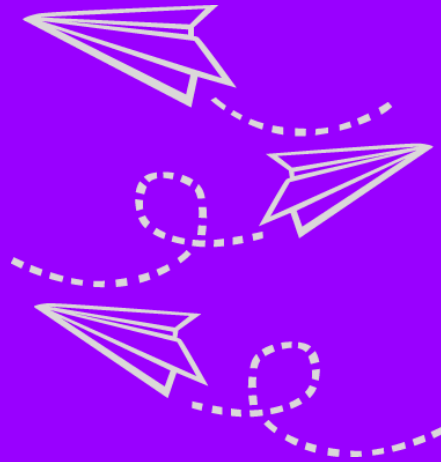
Bibliografía: "An Introduction to Generalized Linear Models". New York: Chapman and Hall, 1990.

CONCLUSIONES

Una visión más moderna y humanizada para atraer a los consumidores digitales.

Es la oportunidad de crear organizaciones y marcas “con alma”, que sean queridas por sus clientes y la sociedad en general





FRASE

«"Con la recopilación de datos, 'cuanto antes mejor' es siempre la mejor respuesta. "»
– Marissa Mayer

«" Para llamar en el estadístico después del experimento se lleva a cabo puede ser no más que pedirle que realice un examen post-mortem:él es capaz de decir, de qué experimento murió él. "». – Ronald Fisher





INICIO RECESO



FIN DE RECESO



FUNDACIÓN DE EDUCACIÓN SUPERIOR

SAN JOSÉ

INSTITUCIÓN TECNOLÓGICA

FIN DE
GRABACIÓN