

# **Universidad de Costa Rica**

## **Informática Aplicada a los Negocios**

### **Investigación: Minería de datos**

#### **Profesora:**

Rolando de Jesús Herrera Sánchez.

#### **Estudiantes**

Steven Naranjo B55008

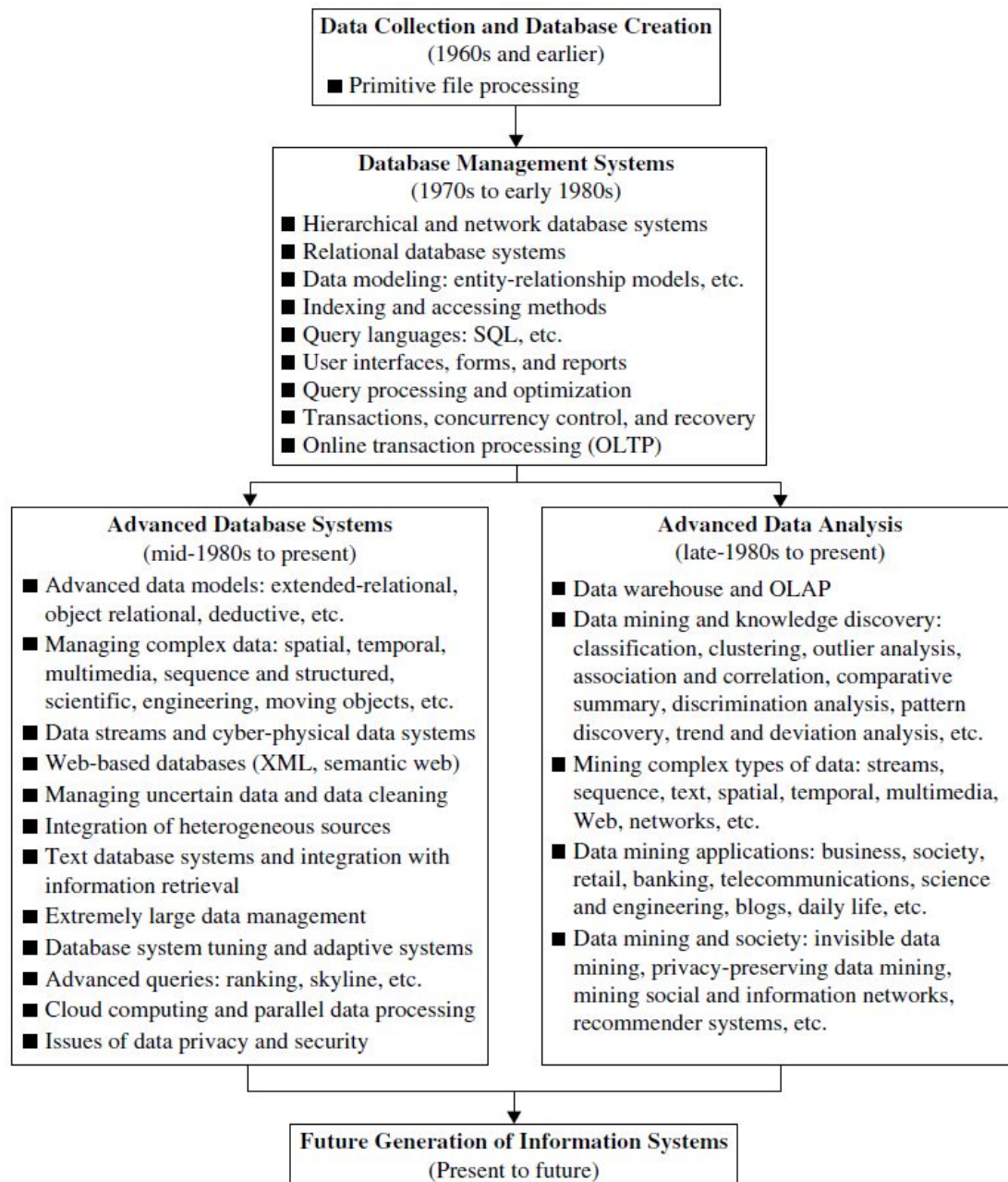
Jasson Romero B76828

Luis Brenes B71300

<b>Minería de datos</b>	<b>4</b>
<b>Historia de la minería de datos</b>	<b>5</b>
<b>¿Qué es la minería de datos?</b>	<b>7</b>
<b>Descubrimiento de conocimiento a partir de los datos (KDD):</b>	<b>7</b>
<b>¿Qué tipo de datos se pueden extraer?</b>	<b>9</b>
Bases de datos	9
Almacenes de datos	9
Datos transaccionales	10
Otros tipos de datos	10
<b>Proceso de desarrollo del modelo de minería de datos.</b>	<b>11</b>
<b>Técnicas aplicadas para la extracción de datos.</b>	<b>13</b>
Técnicas de Clusterizado:	14
K-Means	14
Redes SOM (Self Organization Maps)	15
Sistema de clasificación de la automatización Bayesiana (Autoclass)	15
Clasificadores:	15
Generadores de árboles de decisión.	16
Clasificadores fuzz.	16
Técnicas de Reducción de Dimensión y Visualización de la Información:	16
Análisis de Componentes Principales (PCA)	16
Proyección Pursuit:	16
<b>Herramientas de extracción de datos</b>	<b>17</b>
Clementine de SPSS	17
WEKA	17
SAS Enterprise Miner	18
<b>Referencias</b>	<b>21</b>

## **Minería de datos**

En algún momento hemos escuchado el dicho “*Vivimos en la era de la informática*” pero se podría decir que en realidad vivimos en una era de datos donde día con día surgen grandes cantidades de datos sobre diferentes aspectos de la vida. Todo esto gracias a los grandes avances de la tecnología que ha permitido desarrollar poderosas herramientas para la recopilación y almacenamiento de datos. Por lo cual en este documento se buscará explicar de la mejor manera diferentes aspectos de la minería de datos la cual da uso a todos estos datos generados diariamente en el mundo.



## Historia de la minería de datos

Se dice que desde la década de 1960 la tecnología de la información han estado en continua evolución hasta llegar a lo que se conoce hoy en día, Han, J., Pei, J., & Kamber, M. (2011) mencionan que “la tecnología de la información y las bases de datos ha evolucionado sistemáticamente desde sistemas primitivos de procesamiento de archivos hasta sistemas de bases de datos potentes y sofisticados”(p.4). Por otra parte, en años posteriores a los de 1960 se continuaron dando pasos muy importantes en los temas de tecnología de información

como el hecho que surgen los lenguajes de consulta lo cual facilita mucho el acceso de la información.

La investigación y el desarrollo en sistemas de bases de datos desde la década de 1970 progreso desde los primeros sistemas de bases de datos jerárquicos y de red hasta los sistemas de bases de datos relacionales, herramientas de modelado de datos y métodos de indexación y acceso (Han et al, p.4, 2011).

Con estos avances que se dan en esta década se da también el establecimientos de los sistemas de administración de bases de datos con lo cual, “la tecnología de bases de datos avanzó hacia el desarrollo de sistemas de bases de datos avanzados, almacenamiento de datos y minería de datos para análisis de datos avanzados y bases de datos basadas en la web.” (Han et al, p.4, 2011). Posterior, en la década de 1980 se le inició a dar una mayor importancia al análisis de datos avanzados esto porque en estos años surgen grandes avances en el hardware de las computadoras lo que facilita muchas de las tareas necesarias para realizar dicho análisis dentro de ellas la aparición de diferentes repositorios y tipos de bases de datos. Han et al. (2011) menciona que en estos años:

Una arquitectura de repositorio de datos emergente es el almacén de datos. Este es un repositorio de múltiples fuentes de datos heterogéneas organizadas bajo un esquema unificado en un solo sitio para facilitar la toma de decisiones de gestión. La tecnología de almacenamiento de datos incluye limpieza de datos, integración de datos y procesamiento analítico en línea (OLAP), es decir, técnicas de análisis con funcionalidades como resumen, consolidación y agregación, así como la capacidad de ver información desde diferentes ángulos (p.4).

En la década de 1990 inicia la aparición de World Wide Web (sistema de distribución de documentos de hipertexto) y las bases de datos basadas en web por lo cual ya había mucha información acumulada más allá de las bases de datos o algún almacén de datos por ende “la abundancia de datos, junto con la necesidad de poderosas herramientas de análisis de datos, ha sido descrita como una situación rica en datos pero pobre en información” (Han et al, p.5, 2011). Por lo cual hay muchos datos pero no existe la manera de procesar todo, se hacen intentos para desarrollar sistemas que permitan el procesamiento de todos estos datos pero no se han llegado a una solución concreta a lo cual Han et al. (2011) menciona “La brecha cada vez mayor entre los datos y la información requiere el desarrollo sistemático de herramientas

de minería de datos que puedan convertir las tumbas de datos en "pepitas de oro" de conocimiento" (p.5).

### **¿Qué es la minería de datos?**

La Minería de Datos se podría decir que es un proceso en el cual se intenta descubrir modelos de grandes cantidades de datos. Básicamente consiste en extraer información de un conjunto grande de datos, para luego transformarlos en una estructura que sea comprensible para su uso posterior. Riquelme Santos, J. C., Ruiz, R., & Gilbert, K. (2006) afirman "La Minería de Datos se podría abstraer como la construcción de un modelo que ajustado a unos datos proporciona un conocimiento." (p.15). A su vez esto se puede confirmar según lo dicho por Han et al. (2011) menciona que "la minería de datos debería haberse denominado más apropiadamente *minería de conocimiento a partir de datos*" (p.6).

### **Descubrimiento de conocimiento a partir de los datos (KDD):**

La minería de datos usualmente es confundida con el término "descubrimiento de conocimiento a partir de los datos" aunque también hay personas que dicen que la minería de datos más bien es uno de los pasos en el proceso del descubrimiento de conocimiento a partir de datos, Han et al.(2011) menciona que los pasos a seguir en este proceso son los siguientes.

- “ 1. Limpieza de datos (para eliminar el ruido y los datos incoherentes)
2. Integración de datos (donde se pueden combinar varias fuentes de datos)
3. Selección de datos (cuando los datos relevantes para la tarea de análisis se recuperan del base de datos)
4. Transformación de datos (donde los datos se transforman y consolidan en formularios apropiado para la minería mediante la realización de operaciones de resumen o agregación)
5. Minería de datos (un proceso esencial en el que 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 se utilizan técnicas de visualización y representación del conocimiento para presentar el conocimiento extraído a los usuarios)” (pp.6-8).

Para iniciar el proceso primero tenemos la limpieza de datos la cual viene siendo uno de los procesos más importantes esto debido a que se van a excluir todos aquellos datos que no tengan coherencia alguna con el objetivo que se está planteando. Seguidamente se presenta la integración de datos la cual permite que se pueda realizar una mezcla de datos de varias fuentes lo cual va a ser de mucha utilidad debido a que a veces se presentan datos que son importantes pero incompletos y con esta mezcla de datos se podría conseguir una información que sea muy valiosa.

Seguidamente en este proceso que se lleva a cabo para el descubrimiento de conocimiento a partir de datos, se presenta la selección de datos la cual consta en seleccionar los datos con más relevancia para el proyecto o objetivo que se esté buscando solucionar, y poder realizar el análisis sobre lo que realmente se va a necesitar. Posteriormente se aplica una transformación de todos los datos la cual permite que se puedan ordenar de la manera más eficiente posible, entre estas se mencionan la creación de formularios y resúmenes, los cuales van a permitir que se realice de la manera más apropiada y efectiva posible.

Como quinto paso tenemos la minería de datos la cual es un proceso esencial para el descubrimiento de información, en esta se puede decir que se aplican distintos procedimientos inteligentes los cuales permiten extraer los patrones de datos que son más relevantes para la actualidad. Luego se realiza un análisis sobre estos patrones esto con el fin de obtener los que realmente son de utilidad, con el interés que se tenga detrás de la búsqueda que se esté realizando. Por último se realiza la presentación del conocimiento adquirido, esto lo que va a permitir es mostrar todos los datos extraídos ya con un respectivo orden y una respectiva finalidad. Lo que busca es poder mostrar a los diferentes espectadores lo que se puede llegar a lograr si se sigue este procedimiento, la eficacia que puede llegar a tener el recolectar datos de esta manera de la mano todo con la minería de datos.

Gracias a este proceso se puede llegar a conseguir grandes cosas, la minería de datos de la mano de este proceso ha conseguido ser indispensable en la actualidad y lo seguirá siendo mientras exista ahí afuera tantos datos los cuales sea necesario procesarla para hacer de esto información útil para las personas.

### **¿Qué tipo de datos se pueden extraer?**

La minería se puede implementar sobre cualquier tipo de dato siempre y cuando dicho dato sea significativo para alguien. ¿Pero entonces cuáles serían las principales formas de datos a los cuales se aplica la minería de datos? Han et al. (2011) proponen que “las formas más básicas de datos para aplicaciones de minería son datos de bases de datos, datos de almacenes de datos y datos transaccionales” (p.8). Pero no es que únicamente se pueda aplicar a estas formas de datos ya que como se había mencionado la minería de datos se puede aplicar a cualquier dato por ende Han et al. (2011) menciona que hay otras formas de datos a los cuales se les puede aplicar la minería, estas serían “flujos de datos, datos ordenados / secuenciados, datos de gráficos o en red, datos espaciales, datos de texto, datos multimedia y la WWW” (p.8).

### **Bases de datos**

Un sistema de bases de datos o también llamado sistema de administración de bases de datos según se puede considerar como “una colección de datos interrelacionados, conocida como base de datos, y un conjunto de programas de software para administrar y acceder a los datos” (Han et al, p.9, 2011).

### **Almacenes de datos**

Los almacenes de datos son una de las formas de datos a los cuales se les puede aplicar la minería de datos. ¿Pero qué es un almacén de datos exactamente? Según Han et al. (2011) sería lo siguiente:

Un almacén de datos es un depósito de información recopilada de múltiples fuentes, almacenada bajo un esquema unificado y que generalmente reside en un solo sitio. Los almacenes de datos se construyen mediante un proceso de limpieza de datos, integración de datos, transformación de datos, carga de datos y actualización periódica de datos (p.10).

Los almacenes de datos son de uso muy importante para las empresas que son muy grandes y cuentan con diferentes bases de datos en muchos lugares, esto porque si dicha



empresa quisiera procesar toda esta información para tener un panorama general de todo lo mas facil seria tener toda esta almacenada en un mismo lugar y aquí es donde se daría el uso de los almacenes de datos.

### **Datos transaccionales**

Los datos transaccionales son una fuente de información muy importante para una empresa esto porque ayuda en la toma de decisiones, Han et al. (2011) menciona que “cada registro de una base de datos de transacciones captura una transacción, como la compra de un cliente, la reserva de un vuelo o los clics de un usuario en una página web” (p.13). Esto se vuelve de mucho valor porque se puede tener una idea de usualmente como un usuario usa algún sistema pero, además “una base de datos transaccional puede tener tablas adicionales, que contienen otra información relacionada con las transacciones” (Han et al, p.13, 2011). En estas tablas para ser más específicos se puede almacenar información de productos, sucursales o vendedores.

### **Otros tipos de datos**

Otros de los datos que se presentan a la hora de realizar la minería de datos, son igual de importantes que todos sobre los que se ha hecho mención anteriormente, pero quizás son menos comunes en este proceso. Tal como lo menciona Han et al. (2011)

Existen muchos otros tipos de datos que tienen formas y estructuras versátiles y bastante diferentes significados semánticos. Este tipo de datos se pueden ver en muchas aplicaciones: relacionados con el tiempo o secuencia de datos (por ejemplo, registros históricos, datos de la bolsa de valores y datos de series de tiempo y secuencia biológica), flujos de datos (por ejemplo, videovigilancia y datos de sensores, que son transmisión continua), datos espaciales (por ejemplo, mapas), datos de diseño de ingeniería (por ejemplo, el diseño de edificios, componentes del sistema o circuitos integrados). datos de hipertexto y multimedia (incluidos datos de texto, imágenes, video y audio).(p.14).

Como se puede apreciar los datos tienen mucha importancia ya que entre estos se mencionan los registros históricos, los cuales forman lo que somos hoy en día, también hace mención a la bolsa de valores y secuencia biológica los cuales son datos sensibles que también hace uso la minería de datos por lo cual es importante recalcar la gran área que abarca esta.

### **Proceso de desarrollo del modelo de minería de datos.**

Como bien se ha comentado la minería de datos es la encargada de recolectar y procesar toda la información que sea importante de un gran conjunto de datos, se hace mención que la minería de dato utiliza el análisis matemático para lograr deducir la tendencia y los patrones que hay entre los datos esto según lo que menciona la página oficial de microsoft hacen mención de que si la minería de datos prácticamente sería imposible conseguir esos patrones esto debido a la gran cantidad de datos que o a que se porque la relación entre los mismos es muy compleja. Cabe recalcar que estos patrones los hacen llamar modelos de la minería de datos y mencionan los escenarios en los cuales estos tendrían lugar:

1. Se pueden generar recomendaciones a un cliente, de los productos que puede comprar en un mismo conjunto, según las compras hechas por otros usuarios.
2. Análisis de los artículos que los clientes han introducido en el carrito de la compra y predicción de posibles eventos.
3. Distribución de clientes o eventos en grupos de elementos relacionados, y análisis y predicción de afinidades. (Microsoft, 2009, parr. 5-6-7).

Estos son algunos de los patrones que se logran obtener a través de la minería de datos, los cuales como se puede ver pueden llegar a ser una ventaja competitiva para una empresa, solo el hecho de que un usuario entre a un sitio web y se le presente un conjunto de productos que le pueden llegar a funcionar de una mejor manera juntos, cabe la posibilidad de que el usuario elija esa opción y compre quizás más de lo que pensaba, beneficiando directamente a la empresa.

Seguidamente se presentan cada uno de los pasos del proceso que se realiza para la creación de un modelo de minería de datos, Microsoft. (2009).

1. Definir el problema

2. Preparar los datos
3. Explorar los datos
4. Crear modelos
5. Explorar y validar modelos
6. Implementar y actualizar los modelos (parr.9).

Para iniciar el proceso de minería de datos lo primero que se debe de tener bien claro es cuál es el problema que esto va a resolver, hay que definir el problema de la mejor manera esto para poder encontrar la tendencia y los patrones que tienen los datos que se van a analizar, esto con el fin de encontrar los datos que van a resolver este problema. Seguidamente se debe de realizar un proceso en cual se limpia y consolidan los datos que se obtuvieron en el paso anterior, esto debido a que los datos pueden estar desordenados o bien almacenados en un formato distinto lo cual primero se debe de corregir antes de seguir con el proceso de minería.

El tercer paso de este proceso consiste en realizar un análisis sobre todos los datos que se obtuvieron y están listos para el proceso, esto debido a que al crear los modelos de datos se debe de tener muy claro sobre que se está trabajando para poder tomar las mejores decisiones para la creación de los mismos, se menciona que uno de los cálculos importantes al realizar este proceso es, calcular los valores mínimos y máximos así como calcular la media y las desviaciones estándar.

El siguiente paso es la creación de modelos los cuales van a proceder a crearse con la información o el conocimiento adquirido en el proceso de exploración de los datos, para ello deberá crear columnas de datos que desea utilizar. La función de un modelo de minería de datos es definir cuales son las columnas que se van a utilizar para la entrada de datos así como los parámetros que le van a decir a un determinado algoritmo cual es la mejor manera de procesar los datos que fueron seleccionados para este proceso.

Como quinto paso tenemos el análisis de los modelos que fueron creados en el paso anterior, esto con el fin de determinar si los mismos logran cumplir su objetivo y si realmente lo que se hizo va a producir resultados que sean eficientes para una empresa por ejemplo. Básicamente los modelos van a sufrir un proceso en el cual se van a probar y se va decidir si funcionan correctamente antes de que salgan a producción.

Y por último se presenta la implementación y actualización de estos modelos, el principal objetivo de este paso consiste en implementar los modelos que tienen una mayor

eficiencia en un entorno de producción real, una vez ahí cada uno de los modelos va a realizar distintas tareas las cuales van a permitir poder extraer patrones de la mejor manera, así como muchas otras tareas que van a tener que realizar dependiendo de sus necesidades. Un ejemplo de las tareas que podría realizar uno de los modelos es crear predicciones que luego podrán usar para tomar decisiones comerciales.

### **Técnicas aplicadas para la extracción de datos.**

Como se vio anteriormente se puede decir que la minería de datos es un conjunto de metodologías y herramientas que mediante el análisis de grandes cantidades de datos nos ayudan a obtener patrones de comportamiento o tendencias ocultas que pueden ser muy útiles en la toma de decisiones. Es en este punto importante determinar cuales son las técnicas, metodologías y herramientas que se utilizan para la extracción de estos datos.

Pisón, J y compañía (2002) aclara que la minería de datos, MD de ahora en adelante, no se basa en una metodología estándar y genérica, la cual sirva en todas sus aplicaciones con el mismo formato. Por el contrario es una metodología dinámica e iterativa que se va a ver altamente afectada por 5 aspectos:

1. Problema planteado. No en todas las implementaciones se busca resolver el mismo problema, y como veremos posteriormente, la MD se puede utilizar en diferentes áreas del conocimiento, por esta razón el problema va a tener un amplio abanico de tipos.
2. Disponibilidad de las fuentes de datos
3. Conocimiento de herramientas necesarias. Según la solución que se busca se van a utilizar diferentes tipos de herramientas. Adelante se explicaran algunas herramientas de extracción de datos existentes en el mercado.
4. Metodología desarrollada. Es importante conocer los diferentes tipos de metodologías de MD, para optimizar los resultados de toda la operación en general. De esto Riquelme y compañía destacan el siguiente ejemplo concreto: “no sería apropiado aplicar regresión a unos datos constituidos por texto o modelos basados en distancia a datos simbólicos.”
5. Requerimiento y recursos de la empresa. Al igual que con el problema, todas las empresas van a tener requerimientos y recursos diversos.

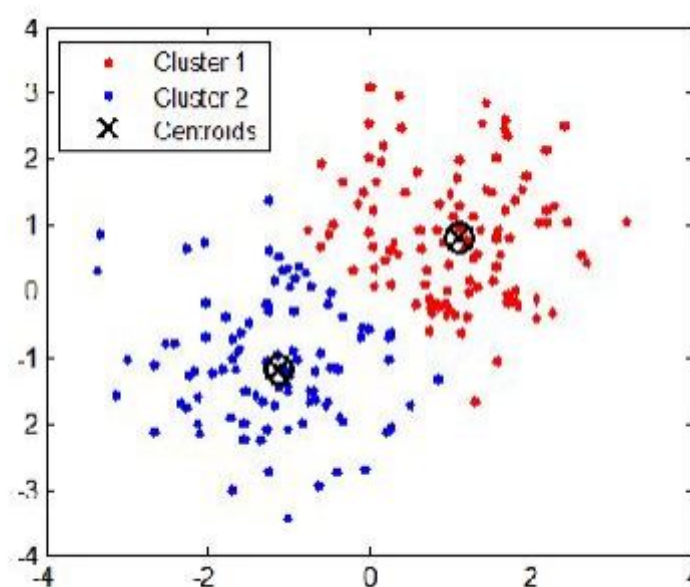
Es importante tomar en cuenta estos 5 factores ya que pueden alterar la adaptabilidad de las técnicas en cada caso de aplicación. Dicho lo anterior Wang Z (1999), categoriza las técnicas de búsqueda de patrones o tendencias ocultas en 3 grandes categorías:

## **I. Técnicas de Clusterizado:**

Se basa en observaciones, las cuales están alojadas en bases de datos. Estas observaciones se clasifican siguiendo identificadores que las caracterizan, por lo cual terminan agrupadas con otras que sean similares. De esta categoría las técnicas más utilizadas son las siguientes:

### **A. K-Means**

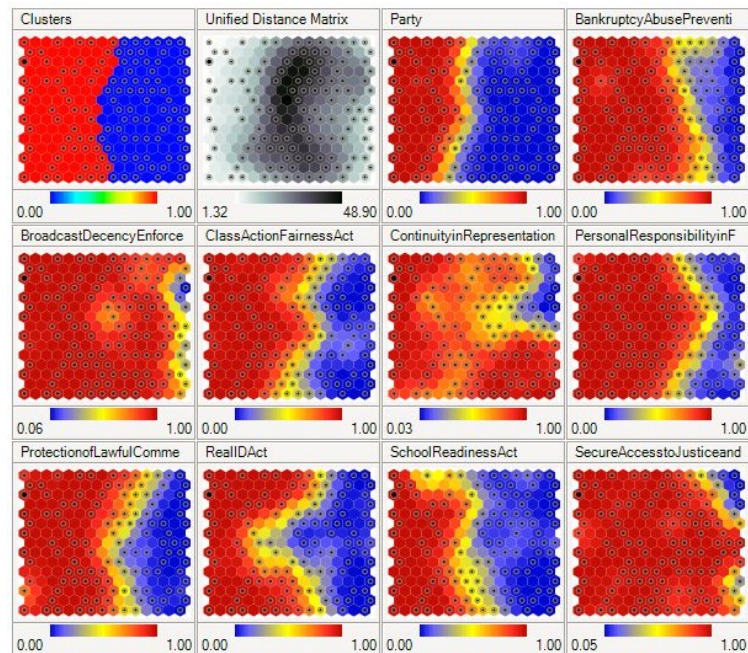
Sigue un procedimiento simple de clasificación de un conjunto de objetos en un determinado número K de clusters. Según Cambronero CG el nombre de K-means representa cada uno de los clusters por la media (o media ponderada) de sus puntos, La siguiente imagen representa gráficamente esta definición.



### **B. Redes SOM (Self Organization Maps)**

El mapa auto-organizado de características es un tipo de red neuronal artificial que produce una representación espacial de las muestras de

entrada, llamado mapa. Tiene como característica que estos usan una función de vecindad para preservar las propiedades. EL siguiente es un ejemplo que muestra los patrones de votación del Congreso de EUA tomado como experimento.



### C. Sistema de clasificación de la automatización Bayesiana (Autoclass)

Es un modelo gráfico probabilístico que representa un conjunto de variables aleatorias y sus dependencias condicionales a través de un grafo acíclico dirigido.

## II. Clasificadores:

Pisón, J y compañía (2002) define esta categoría como aquellos patrones que explican el comportamiento de unas variables frente a otras. Se dividen fundamentalmente en:

### A. Generadores de árboles de decisión.

Un árbol de decisión es un modelo de predicción utilizado en la MD en la cual se fabrican diagramas de construcciones lógicas, que sirven para representar y categorizar datos y observaciones, para la resolución de un problema.

B. Clasificadores fuzz.

Según Piorno J, consiste en dividir primero el conjunto de muestras en clases, luego aparece una fase de optimización para reducir el número de parámetros utilizados en cada clase, y finalmente combina las clases locales obteniendo una solución final que encuentra una

**III. Técnicas de Reducción de Dimensión y Visualización de la Información:**

Esta serie de técnicas tiene por objetivo reducir el número de variables y visualizar la nube de puntos N-dimensionales para poder detectar estructuras o características de forma visual. Sus principales técnicas son:

A. Análisis de Componentes Principales (PCA)

Es una técnica utilizada para describir un conjunto de datos en términos de nuevas variables no correlacionadas. ACP se emplea sobre todo en análisis exploratorio de datos y para construir modelos predictivos

B. Proyección Pursuit:

Técnica estadística utilizada para análisis exploratorio de datos , visualización de la información , y la selección de características . Permite al usuario explorar de forma interactiva datos muy complejos para encontrar características o patrones de interés potencial.

Como última consideración, la elección de una técnica, sin importar cual sea, hay que considerar dos condicionantes: el tipo de los datos y el objetivo que se quiera obtener.

**Herramientas de extracción de datos**

Para la extracción de datos se requieren una serie de herramientas que permiten analizar datos y transformarlos en información útil para la toma de decisiones. Ante esto Yuniet R alcara que las herramientas se pueden clasificar en dos grandes categorías:

Herramientas de verificación: se limitan a comprobar hipótesis suministrada por los usuarios.

Herramientas de descubrimiento: Todas aquellas que se encargan de encontrar patrones potencialmente interesantes de forma automática, incluyendo en este grupo todas las técnicas de predicción.

Dicho lo anterior, a continuación se le presentan las principales herramientas que se encuentran en el mercado para extraer conocimientos desde bases de datos. Suárez R., y Amador D. (2009) da los siguientes ejemplos:

- Clementine de SPSS

Es una aplicación de software de análisis de texto y minería de datos de IBM . Se utiliza para construir modelos predictivos y realizar otras tareas analíticas. Clementine principalmente permite:

- Acceder, preparar e integrar fácilmente datos numéricos, de texto, datos provenientes de páginas Web y de encuestas.
- Construir y validar modelos rápidamente utilizando las técnicas estadísticas y de aprendizaje automático disponibles más avanzadas
- Obtener rápidamente un mejor retorno de la inversión y mejores tiempos de respuesta aprovechando las características de rendimiento y escalabilidad.

- WEKA

Desarrollada en Java, de código abierto y con licencia GPL, Weka proporciona gran cantidad de herramientas para la realización de tareas propias de minería de datos, la visualización y permite la programación en Java de algoritmos más sofisticados para análisis de datos y modelado predictivo.

- SAS Enterprise Miner

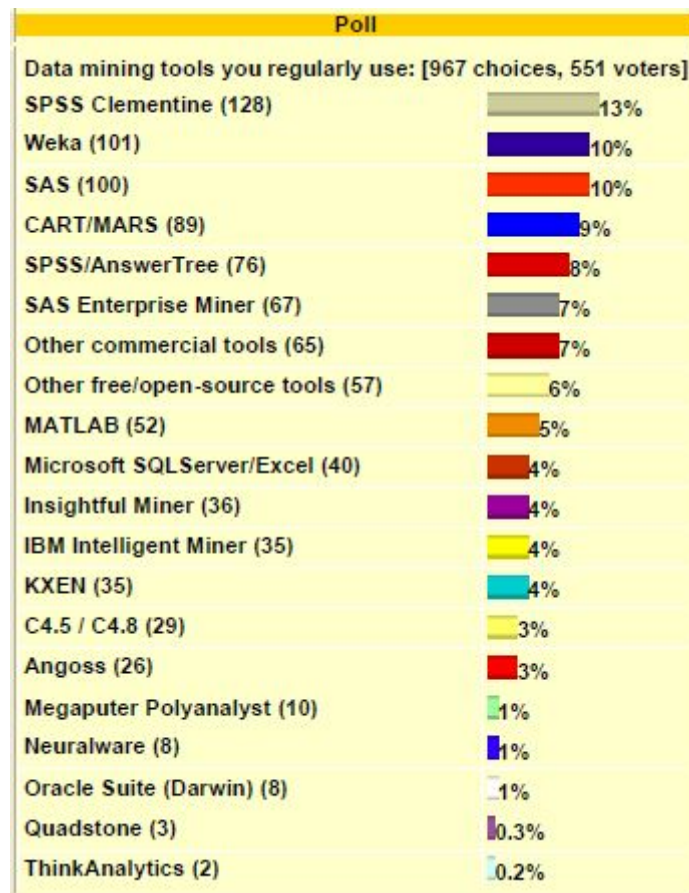
Solución de minería de datos que permite incorporar patrones inteligentes a los procesos de marketing, tanto operativos como estratégicos. Dentro de las principales características encontramos:

- Incluye su propia base de datos de información para almacenar y manejar los datos, es decir, un "data warehouse".



- Soporta los principales protocolos de comunicación,
- Cubre los cinco modelos de procesamiento cliente/servidor de acuerdo a Gartner Group
- Cumple con las 12 reglas de OLAP

Así como estos tres ejemplos existen muchos más, y Pisón, J y compañía (2002) destacan en la siguiente imagen, en la cual se puede apreciar las principales herramientas utilizadas la minería de datos según la encuesta hecha en el conocido portal sobre Minería de Datos y Gestión del Conocimiento, KDnuggets.



### Usos de la minería de datos

Una de las principales tareas que tienen la minería de datos es reconocer cual es la utilidad de las técnicas existentes. Está también pretende ampliar nuevas técnicas que permitan una mejora en el mercado tal como Riquelme , Ruiz & Gilbert (2006). Afirman “La

minería de datos busca desarrollar nuevas 1 para dominios tradicionales o de nueva aplicación, como el comercio electrónico y la bioinformática” (p.14).

Por lo cual la MD tiene un aporte muy importante en cuanto a la ayuda que le pretende brindar a las áreas que generan datos, seguidamente se mencionan algunas de las áreas en las cuales es más común que se presente la minería de datos. Riquelme , Ruiz & Gilbert (2006) brinda los siguientes ejemplos:

- Comercio y banca. La minería de datos puede contribuir significativamente en las aplicaciones de administración empresarial basada en la relación con el cliente. Dentro las sus mayores aportes se encuentra:
  - Segmentación del mercado. Analizando las relaciones entre parámetros como edad de los clientes, género, gustos, es posible adivinar su comportamiento para dirigir campañas personalizadas de fidelización o captación
  - Previsión de ventas. Detectar patrones de compra conjunta para identificar asociaciones de productos.
  - Análisis de riesgo. Es habitual que se aplique a la calificación crediticia y a sistemas inteligentes antifraude para analizar transacciones, movimientos de tarjetas, patrones de compra y datos financieros de los clientes.
- Medicina y Farmacia: diagnóstico de enfermedades y la efectividad de los tratamientos. Al contar con toda la información del paciente —historial, examen físico y patrones de terapias anteriores— se pueden prescribir tratamientos más efectivos. También posibilita una gestión más eficaz, eficiente y económica de los recursos sanitarios al identificar riesgos, predecir enfermedades en ciertos segmentos de la población o pronosticar la duración del ingreso hospitalario.
- Seguridad y detección de fraude: reconocimiento facial, identificaciones biométricas, accesos a redes no permitidos, etc. Generalmente, estas operaciones fraudulentas o ilegales suelen seguir patrones característicos que permiten, con cierto grado de probabilidad, distinguirlas de las legítimas y desarrollar así mecanismos para tomar medidas rápidas frente a ellas
- Recuperación de información no numérica: minería de texto, minería web, búsqueda e identificación de imagen, video, voz y texto de bases de datos multimedia.
- Astronomía: identificación de nuevas estrellas y galaxias.

- Geología, minería, agricultura y pesca: identificación de áreas de uso para distintos cultivos o de pesca o de explotación minera en bases de datos de imágenes de satélites
- Ciencias Ambientales: identificación de modelos de funcionamiento de ecosistemas naturales y/o artificiales (p.e. plantas depuradoras de aguas residuales), para mejorar su observación, gestión y/o control.
- Ciencias Sociales: Estudio de los flujos de la opinión pública. Planificación de ciudades: identificar barrios con conflicto en función de valores sociodemográficos.

## **Referencias**

Argosy Medical Animation. (2019). Visible body: Discover human anatomy. New York, EU.: Argosy Publishing. Recuperado de <http://www.visiblebody.com>

Cambronero, CG y Moreno, IG (2006). Algoritmos de aprendizaje: knn & kmeans. *Inteligencia en Redes de Comunicación, Universidad Carlos III de Madrid* , 23

Hand, D. J., & Adams, N. M. (2014). *Data Mining. Wiley StatsRef: Statistics Reference Online*, 1-7. Recuperado de <https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118445112.stat06466.pub2>

Han, J., Pei, J., & Kamber, M. (2011). *Data mining: concepts and techniques*. Elsevier.

Martínez de Pisón, F. J., Pernía Espinoza, A., Castejón Limas, M., & González Marcos, A. (2002). Minería de Datos: Herramientas, Técnicas y Metodologías. In *VI International Congress on Project Engineering. Barcelona*. Recuperado de

[https://www.aepro.com/files/congresos/2002barcelona/ciip02\\_0359\\_0367.1895.pdf](https://www.aepro.com/files/congresos/2002barcelona/ciip02_0359_0367.1895.pdf)

Microsoft. (2019). Conceptos de minería de datos. Lugar de publicación: *microsoft.com*  
Recuperado de:  
<https://docs.microsoft.com/es-es/analysis-services/data-mining/data-mining-concepts?view=asallproducts-allversions#DefiningTheProblem>

*Pujari, A. K. (2001). Data mining techniques. Universities press.*

Riquelme Santos, J. C., Ruiz, R., Gilbert, K. (2006). Minería de datos: Conceptos y tendencias. *Inteligencia Artificial: Revista Iberoamericana de Inteligencia Artificial*, 10 (29), 11-18. Recuperado de <https://idus.us.es/bitstream/handle/11441/43290/Miner%C3%ADa%20de%20datos.pdf?sequence=1&isAllowed=y>

Suárez, Y. R., & Amador, A. D. (2009). Herramientas de minería de datos. *Revista Cubana de Ciencias Informáticas*, 3(3-4), 73-80. Recuperado de <https://www.redalyc.org/pdf/3783/378343637009.pdf>

WANG, XUE Z. (1999). Data Mining and Knowledge Discovery for Process. London: Springer-Verlag.