

## PROGRAMA DEL DIPLOMADO DE PROCESO BENCHMARKING.

### TEMA 7. MANEJO DE LA INFORMACIÓN.

#### **Objetivo:**

Al final de la unidad el alumno comprenderá la presencia de estas herramientas informáticas (programas – Datamining)) necesarias para complementar el registro y control del CRM y serán de un gran apoyo para la toma de decisiones y aplicación de estrategias mercadológicas para la dirección de la empresa.

#### **7.1. Datamining (MINERÍA DE DATOS)**

Muchas organizaciones han acumulado una enorme cantidad de datos en lo que llamamos **Bases de Datos**, cuando lo que realmente necesitan *es información que les ayude a definir potenciales clientes y mercados, maneras de entregarles un mejor servicio a los mismos, en definitiva, a agregar valor a sus procesos.*

Una de las maneras de lograr estos objetivos (que desde luego no son triviales) es la Datamining -minería de datos-. **La minería de datos usa estadística avanzada y algoritmos de inteligencia artificial para descubrir patrones y relaciones escondidas.** La clave para lograr el objetivo planteado es la Minería de Datos, *con esta herramienta se puede extraer conocimiento enfocado a grupos de clientes (con objetivos comunes por ejemplo).* Las bases de datos (a través de la minería de datos) se exploran a través de miles de “puntos de vista”; lo que permite encontrar información escondida acerca del comportamiento de clientes para establecer patrones. Así los departamentos de marketing se adaptan rápidamente a la flexibilidad y los requerimientos de las necesidades de cada cliente que alcanzan.

#### **7.1.1.- Definición:**

Datamining (Minería de datos) es la extracción de información oculta y predecible de grandes bases de datos, es una poderosa tecnología nueva con gran potencial para ayudar a las compañías a concentrarse en la información más importante de sus Bases de Información (DataWarehouse).

La minería de datos utiliza técnicas de estadística y “aprendizaje de las máquinas” para construir modelos que predigan el comportamiento de futuro de los consumidores. Hoy, la tecnología automatiza el proceso de “minería”, la integra a los ya difundidos “data Warehouse” y la presenta como una herramienta relevante y útil para la organización actual. La minería de datos consiste en descubrir información interesante dentro de las bases de datos existentes en la organización pero que se encuentra escondida debido a la gran cantidad de datos almacenados. La minería de datos no reemplaza a la estadística tradicional; es más bien una extensión de la misma, que forma parte de un resultado que ha sido formado por la comunidad estadística cambiando de manera radical ciertos aspectos.

El incremento en el poder de las computadoras y los bajos costos, unidos a la necesidad de analizar enormes bases de datos que contienen millones de filas, han permitido el desarrollo de nuevas técnicas basadas en la exploración a “fuerza bruta” de las posibles soluciones. El punto clave es que la minería de datos es la aplicación de éstas y otras técnicas estadísticas de inteligencia artificial a problemas de negocios en una manera tal, que son alcanzables tanto a usuarios de negocios como para expertos estadistas.

### **7.1.2.- ¿Cómo Trabaja el Datamining?:**

¿Cuán exactamente es capaz Datamining de decirle cosas importantes que usted desconoce o que van a pasar?

La técnica usada para realizar estas hazañas en Datamining se llama Modelado. El modelado es simplemente el acto de construir un modelo en una situación donde usted conoce la respuesta y luego la aplica en otra situación de la cual desconoce la respuesta.

Este acto de construcción de un modelo es algo que la gente ha estado haciendo desde hace mucho tiempo, seguramente desde antes del auge de las computadoras y de la tecnología de Datamining. Lo que ocurre en las computadoras, no es muy diferente de la manera en que la gente construye modelos. Las computadoras son cargadas con mucha información acerca de una variedad de situaciones donde una respuesta es conocida y luego el software de Datamining en la computadora debe correr a través de los datos y distinguir las características de los datos que llevarán al modelo. Una vez que el modelo se construyó, puede ser usado en situaciones similares donde usted no conoce la respuesta.

Si alguien le dice que tiene un modelo que puede predecir el uso de los clientes, ¿Cómo puede saber si es realmente un buen modelo?, la primera cosa que puede probar es pedirle que aplique el modelo a su base de clientes - donde usted ya conoce la respuesta. Con Datamining, la mejor manera para realizar esto es dejando de lado ciertos datos para aislarlos del proceso de Datamining.

Una vez que el proceso está completo, los resultados pueden ser testeados contra los datos excluidos para confirmar la validez del modelo. Si el modelo funciona, las observaciones deben mantenerse para los datos excluidos.

### **7.1.3.- VENTAJAS DE LA MINERÍA DE DATOS:**

En resumen, usar minería de datos para construir un modelo desde las bases de datos tiene las siguientes ventajas.

Los modelos son fáciles de entender. Personas sin un back up importante de estadísticas (como un analista financiero o ejecutivos en general) pueden interpretar el modelo y compararlo con sus propias ideas; recordemos que en la organización de hoy, son los mismos ejecutivos los que deben “meter las manos” a las bases de datos para obtener la información que necesitan. Así, el usuario

gana más conocimientos sobre el comportamiento de los clientes y puede (y debe) usar esta ventaja para optimizar el proceso de negocios de la organización.

Enormes bases de datos pueden ser analizadas. Enormes bases de datos pueden ser analizadas mediante la tecnología de la minería de datos. Estas Bases de datos pueden ser enormes tanto en largo como en ancho. Por ejemplo, para cada cliente se puede tener cientos de atributos que contienen información detallada; y además tener miles de registros de clientes.

La minería de datos descubre información que no se esperaba obtener. Como muchos modelos diferentes son validados, algunos resultados inesperados tienden a aparecer. En muchos estudios, se ha descubierto que combinaciones particulares de factores entregan efectos inesperados que entregan valor a la compañía.

Los Modelos son confiables. El modelo es probado y comprobado usando técnicas estadísticas antes de ser usado, luego las predicciones que se obtienen por el modelo son válidas y confiables.

Los modelos se construyen de manera rápida. La minería de datos permite construir y generar modelos en sólo uno minutos u horas. El modelado se torna mucho más fácil puesto que muchos algoritmos son probados y sólo el mejor modelo es entregado al usuario.

## **7.2.- Aplicaciones:**

Una aplicación, por ejemplo, es implementar un proceso que genere una muy precisa segmentación de los clientes. Una vez que el motor del Excavador de Datos ha seleccionado un grupo adecuado de segmentos de clientes de su Bodega de Datos (DataWarehouse), el próximo paso será extrapolar los perfiles de los consumidores. Cada vez que llega una nueva cosecha de clientes se aplica un nuevo conjunto de modelos estadísticos y se corre el programa para comparar contra los segmentos existentes o crear otros nuevos. Aquí estamos enfocando la atención a predecir la lealtad de marca, por citar un ejemplo, pero tenemos muchas otras aplicaciones como son:

- Segmentación del mercado
- Tendencias de deserción de clientes
- Descubrimiento de transacciones fraudulentas
- Mercadeo directo
- Mercadeo Interactivo
- Análisis de canasta
- Análisis de tendencias
- Perfiles de clientes
- Focalización de clientes y campañas promocionales

### **7.2.1.- El Mercadeo mediante Excavación de Datos:**

Convierte una plataforma tecnológica en un sistema de información sobre el que se construyen soluciones de negocios. Naturalmente el punto de partida es que las montañas de datos deben ser de oro y no de chatarra. De lo contrario no vale la pena excavar. Se necesitarán equipos de cómputo, sistemas operativos y la infraestructura necesaria para apoyar ese proyecto minero. Luego vendrán las técnicas de inteligencia artificial y de análisis estadístico que permitirán extraer el oro de su mina. El primer paso es evaluar los recursos, objetivos y necesidades. Dependiendo de esto se definirá si tecnológicamente el énfasis será por ejemplo hacia procesos cliente-servidor y de escritorio o hacia procesos en paralelo. En la actualidad, existe una gran cantidad de aplicaciones, en áreas tales como:

- Astronomía: clasificación de cuerpos celestes.
- Aspectos climatológicos: predicción de tormentas, etc.
- Medicina: caracterización y predicción de enfermedades, probabilidad de respuesta satisfactoria a tratamiento médico.
- Industria y manufactura: diagnóstico de fallas.
- Mercadotecnia: identificar clientes susceptibles de responder a ofertas de productos y servicios por correo, fidelidad de clientes, selección de sitios de tiendas, afinidad de productos, etc.
- Inversión en casas de bolsa y banca: análisis de clientes, aprobación de préstamos, determinación de montos de crédito, etc.
- Detección de fraudes y comportamientos inusuales: telefónicos, seguros, en tarjetas de crédito, de evasión fiscal, electricidad, etc.
- Análisis de canastas de mercado para mejorar la organización de tiendas, segmentación de mercado (clustering)
- Determinación de niveles de audiencia de programas televisivos
- Normalización automática de bases de datos

### **7.2.2.- HERRAMIENTAS:**

Las herramientas de datamining o minería de datos pueden responder a preguntas de negocios empresariales a priori no planteadas o que pueden consumir demasiado tiempo para ser resueltas. Los programas de gestión del conocimiento se complementan con distintas herramientas además del datamining, como puede ser el datawarehousing o el groupware. El datamining, como herramienta de búsqueda de información, se utiliza como sistema de apoyo a la toma de decisiones de las altas direcciones de las empresas.

En definitiva, la minería de datos es una tecnología usada para descubrir información oculta y desconocida, pero potencialmente útil, a partir de las fuentes de información de la propia empresa. Obtiene un conocimiento de un negocio, utilizando técnicas de clustering, redes neuronales, árboles de decisión y reglas de asociación etc. Como ya se ha comentado, las técnicas de la minería de datos provienen de la Inteligencia artificial y de la estadística, dichas técnicas, no son más que algoritmos, más o menos sofisticados que se aplican sobre un conjunto de datos para obtener unos resultados.

Las técnicas más representativas son:

- Modelos estadísticos. Es una expresión simbólica en forma de igualdad o ecuación que se emplea en todos los diseños experimentales y en la regresión para indicar los diferentes factores que modifican la variable de respuesta.
- Agrupamiento o Clustering. Es un procedimiento de agrupación de una serie de vectores según criterios habitualmente de distancia; se tratará de disponer los vectores de entrada de forma que estén más cercanos aquellos que tengan características comunes. Ejemplos: Algoritmo K-means, algoritmo K-medoids.

Según el objetivo del análisis de los datos, los algoritmos utilizados se clasifican en supervisados y no supervisados:

- Algoritmos supervisados (o predictivos): predicen un dato (o un conjunto de ellos) desconocido a priori, a partir de otros conocidos.
- Algoritmos no supervisados (o del descubrimiento del conocimiento): se descubren patrones y tendencias en los datos.

### **7.2.3.- Análisis preliminar de datos usando Query tolos:**

El primer paso en un proyecto de datamining sería siempre un análisis de los datos usando query tools, aplicando una consulta SQL a un conjunto de datos, para rescatar algunos aspectos visibles antes de aplicar las técnicas. La gran mayoría de la información (un 80 %) puede obtenerse con SQL. El 20 % restante, más importante, la información oculta requiere técnicas avanzadas. Este primer análisis en SQL es para saber cual es la distribución de los valores posibles de los atributos. Recién después podemos ver la performance del algoritmo correspondiente.

Técnicas de visualización. Estas son buenas para ubicar patrones en un conjunto de datos y puede ser usado al comienzo de un proceso de datamining para tomar un feeling de la calidad del conjunto de datos.

Árbol de Decisión. Son estructuras en forma de árbol que representan conjuntos de decisiones. Estas decisiones generan reglas para la clasificación de un conjunto de datos. Para poder predecir el comportamiento de un cliente es necesario poder contar con una clasificación previa esto implica una predicción de que un cliente pertenece a cierto grupo de clientes. La complejidad es de  $n (\log n)$ .

Algunos Métodos específicos de árboles de decisión incluyen:

- CART Árboles de clasificación y regresión: técnica usada para la clasificación de un conjunto de datos. Provee un conjunto de reglas que se pueden aplicar a un nuevo (sin clasificar) conjunto de datos para predecir cuáles registros darán un cierto resultado.
- Segmenta un conjunto de datos creando 2 divisiones. Requiere menos preparación de datos que CHAID. CHAID Detección de interacción automática de Chi cuadrado: técnica similar a la anterior, pero segmenta un conjunto de datos utilizando tests de chi cuadrado para crear múltiples divisiones.
- Reglas de Asociación. Establece asociaciones en base a los perfiles de los clientes sobre los cuales se está realizando el datamining. Las reglas de

Asociación están siempre definidas sobre atributos binarios. No es muy complicado generar reglas en grandes bases de datos. El problema es que tal algoritmo eventualmente puede dar información que no es relevante. Datamining envuelve modelos para determinar patterns a partir de los datos observados. Los modelos juegan un rol de conocimiento inferido. Diciendo cuando el conocimiento representa conocimiento útil o no, esto es parte del proceso de extracción de conocimiento en bases de datos (Knowledge Discovery in Databases-KDD).

Algoritmos Genéticos. Son técnicas de optimización que usan procesos tales como combinaciones genéticas, mutaciones y selección natural en un diseño basado en los conceptos de evolución.

Redes Bayesianas. Buscan determinar relaciones causales que expliquen un fenómeno en base a los datos contenidos en una base de datos. Se han usado principalmente para realizar predicción.

Procesamiento Analítico en Línea (OLAP). Estas herramientas ofrecen un mayor poder para revisar, graficar y visualizar información multidimensional, en características temporales, espaciales o propias. Se valen de lenguajes menos restringidos y estructurados como lo es SQL. Requieren todavía de una alta participación de un usuario humano, pues son interactivas y requieren la guía del experto.

Redes Neuronales Artificiales. Son modelos predecibles, no lineales que aprenden a través del entrenamiento y semejan la estructura de una red neuronal biológica. Método del vecino más cercano: una técnica que clasifica cada registro en un conjunto de datos basado en una combinación de las clases de k registro/s más similar/es a él en un conjunto de datos históricos. Algunas veces se llama la técnica del vecino k-más cercano.

Regla de inducción. La extracción de reglas if-then de datos basados en significado estadístico.

Para aplicar mejor estas técnicas avanzadas, éstas deben estar totalmente integradas con el datawarehouse así como con herramientas flexibles e interactivas para el análisis de negocios. Varias herramientas de Datamining actualmente operan fuera del datawarehouse, requiriendo pasos extra para extraer, importar y analizar los datos. Además, cuando nuevos conceptos requieren implementación operacional, la integración con el warehouse simplifica la aplicación de los resultados desde Datamining El Datawarehouse analítico resultante puede ser aplicado para mejorar procesos de negocios en toda la organización, en áreas tales como manejo de campañas promocionales, detección de fraudes, lanzamiento de nuevos productos, etc.

El punto de inicio ideal es un datawarehouse que contenga una combinación de datos de seguimiento interno de todos los clientes junto con datos externos de

mercado acerca de la actividad de los competidores. Información histórica sobre potenciales clientes también provee una excelente base para prospecting. Este warehouse puede ser implementado en una variedad de sistemas de bases relacionales y debe ser optimizado para un acceso a los datos flexible y rápido. Las herramientas de Datamining predicen futuras tendencias y comportamientos, permitiendo en los negocios tomar decisiones proactivas y conducidas por un conocimiento acabado de la información (knowledge-driven). Los análisis prospectivos automatizados ofrecidos por un producto así van más allá de los eventos pasados provistos por herramientas retrospectivas típicas de sistemas de soporte de decisión.

Una vez que las herramientas de Datamining fueron implementadas en computadoras cliente servidor de alto performance o de procesamiento paralelo, pueden analizar bases de datos masivas para brindar respuesta a preguntas tales como, "¿Cuáles clientes tienen más probabilidad de responder al próximo mailing promocional, y por qué?" y presentar los resultados en formas de tablas, con gráficos, reportes, texto, hipertexto, etc. Las técnicas de Datamining pueden ser implementadas rápidamente en plataformas ya existentes de software y hardware para acrecentar el valor de las fuentes de información existentes y pueden ser integradas con nuevos productos y sistemas pues son traídas en línea (on-line).

Sin duda alguna que el uso de Datamining:

- Contribuye a la toma de decisiones tácticas y estratégicas proporcionando un sentido automatizado para identificar información clave desde volúmenes de datos generados por procesos tradicionales y de e-Business
- Permite a los usuarios dar prioridad a decisiones y acciones mostrando factores que tienen un mayor en un objetivo, qué segmentos de clientes son desechables y qué unidades de negocio son sobrepasados y por qué.
- Proporciona poderes de decisión a los usuarios del negocio que mejor entienden el problema y el entorno y es capaz de medir las acciones y los resultados de la mejor forma.
- Genera Modelos descriptivos: En un contexto de objetivos definidos en los negocios permite a empresas, sin tener en cuenta la industria o el tamaño, explorar automáticamente, visualizar y comprender los datos e identificar patrones, relaciones y dependencias que impactan en los resultados finales de la cuenta de resultados (tales como el aumento de los ingresos, incremento de los beneficios, contención de costos y gestión de riesgos)
- Genera Modelos predictivos: permite que relaciones no descubiertas e identificadas a través del proceso del Datamining sean expresadas como reglas de negocio o modelos predictivos. Estos outputs pueden comunicarse en formatos tradicionales (presentaciones, informes, información electrónica compartida, embebidos en aplicaciones, etc.) para guiar la estrategia y planificación de la empresa.

### **7.3.- PUNTOS DE CONTACTO CON EL CLIENTE:**

El "Datamining" integra los procesos modernos de Ventas, excepto el "Key Account Management" que representa la versión unitaria (no masiva) del manejo de Cuentas Clave, en lo cual se pueden aplicar algunas técnicas que señalamos a continuación:

#### **7.3.1.- El "Cruce de Productos"**

Esta técnica se basa en el Mercadeo Concéntrico, esto es, en múltiples ofertas alrededor de un mismo cliente. A mayor cantidad de transacciones o relaciones que sostenga una cuenta con nosotros, mayor será la capacidad de la empresa de retenerla con el paso del tiempo. Este "Cruce de Productos" puede ser personal o masivo. Cada uno requiere de la segmentación de la clientela para adaptar la oferta a las necesidades del cliente o grupo de éstos y de la existencia de alguna Matriz PPC que identifique cuáles productos se le han colocado a cuáles clientes (y cuáles no) para facilitar eventuales ofertas.

#### **7.3.2.- El "Networking":**

Cada cliente puede ser una fuente de negocios adicionales, pero existen siempre evidentes "Multiplicadores" que por su naturaleza o poder de convocatoria aglutinan números importantes de clientes potenciales. Esta técnica busca crear redes de cuentas potenciales alrededor de una sola matriz Internamente, en nuestra clientela o "C-Base" existen normalmente varias de éstas y en el mercado tantas más. Por ejemplo: Pueden ser nuestros clientes los empleados de ellos, sus propios clientes, sus proveedores, afiliados, etc.

En el entorno "CRM" se puede evaluar y desarrollar un conjunto de reglas de negocio sobre todos los aspectos de las interacciones de los clientes. Un ejemplo sencillo sería modelar la probabilidad de respuesta ante una solicitud específica de un nuevo producto o servicio. Basándose en estas reglas las campañas de marketing de una empresa pueden tener como objetivo la máxima respuesta para generar un nivel deseado de respuesta, ingresos o beneficios. Otras aplicaciones incluirían:

- Crear modelos de clientes potenciales (para objetivos de marketing y otras iniciativas CRM).
- Analizar la pérdida de clientes (para atención al cliente y reclamaciones).
- Monitorizar los riesgos de pérdida (para scoring de clientes y aprobación de créditos).
- Investigación del fraude (para reducir pérdidas a través de la investigación del fraude).

Sin embargo, como se dijo anteriormente el alcance de la tecnología del Datamining llega mucho más allá del "CRM" e incluye cualquier proceso que lleva la adquisición, interpretación y actualización en los datos (bien de origen interno o externo).



