



**UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO**

FACULTAD DE CIENCIAS

**MINERÍA DE DATOS PARA LA TOMA DE
DECISIONES E INTELIGENCIA DE NEGOCIOS:
APLICACIONES EN LA MERCADOTECNIA**

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

A C T U A R I O

P R E S E N T A:

ROBERTO GARCÍA REYES



**DIRECTOR DE TESIS:
DRA. AMPARO LÓPEZ GAONA
2012**



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

HOJA DE DATOS DEL JURADO

1. Datos del alumno
García
Reyes
Roberto
55 28 50 12 66
Universidad Nacional Autónoma de México
Facultad de Ciencias
Actuaría
098505131
2. Datos del tutor
Dra
Amparo
López
Gaona
3. Datos del sinodal 1
Dra.
Hanna
Jadwiga
Oktaba
4. Datos del sinodal 2
M. en I.
Gerardo
Avilés
Rosas
5. Datos del sinodal 3
M. en C.
Gustavo Arturo
Márquez
Flores
6. Datos del sinodal 4
M. en C.
Cecilia
Pérez
Colín
7. Datos del trabajo escrito
Minería de Datos para la Toma de Decisiones e Inteligencia de Negocios: Aplicaciones
en la Mercadotecnia
119 p.
2012

INDICE	Página
INTRODUCCIÓN	6
I. INTELIGENCIA DE NEGOCIOS	8
I.1 Bases de datos	10
I.2 Modelos matemáticos	13
I.3 Arquitectura de la inteligencia de negocios	14
I.3.1 Fuentes de datos	14
I.3.2 Almacenes de datos	14
I.3.2.1 Data warehouse	14
I.3.2.2 Data marts	15
I.3.3 Exploración de datos	16
I.3.3.1 Análisis univariado	16
I.3.3.2 Análisis bi variado	17
I.3.3.3 Análisis multivariado	20
I.3.4 Minería de datos	22
I.3.5 Optimización	22
I.3.6 Decisiones	23
I.4 Ciclo del análisis de inteligencia de negocios	24
I.4.1 Análisis	24
I.4.2 Penetración	24
I.4.3 Decisión	24
I.4.4 Evaluación	24
I.5 Software	26
I.6 Inteligencia de negocios en tiempo real	27
I.7 Ejemplos de técnicas para integrar la inteligencia de negocios en una empresa	29
I.7.1 Portales empresariales	29
I.7.2 Incrustar las aplicaciones analíticas en las operacionales	30
I.8 Barreras organizacionales a la inteligencia de negocios	32
II. MINERÍA DE DATOS (DATA MINING)	34
II.1 ¿Qué es la minería de datos?	34
II.2 Tipos de datos necesarios para la aplicación de la minería de datos	39
II.2.1 Bases de datos relacionales	39
II.2.2 Bases de datos transaccionales	39

II.2.3 Sistemas de bases de datos avanzados	39
II.2.3.1 Bases de datos orientadas a objetos	40
II.2.3.2 Bases de datos relacional –objeto	40
II.2.4 Bases de datos espaciales	40
II.2.5 Bases de datos temporales y de series de tiempo	40
II.2.6 Bases de datos de texto y multimedia	41
II.2.7 <i>La World Wide Web</i>	41
II.3 Funcionalidad de la minería de datos	42
II.3.1 Descripción de concepto/clase: Caracterización y discriminación	
II.3.2 Análisis de asociación o asociativo	43
II.3.3 Clasificación y predicción	43
II.3.4 Análisis Cluster o de grupos (definición)	43
II.3.5 Análisis Outlier o de partes aisladas	43
II.3.6 Análisis de evolución	
II.4 El proceso de la minería de datos	45
II.4.1 El descubrimiento y la explotación	45
II.4.2 Modelos	45
II.5 Las matemáticas en la minería de datos	48
II.6 Técnicas de minería de datos	51
II.6.1 Análisis Cluster (explicación)	51
II.6.1.1 Matriz de datos	51
II.6.1.2 Matriz de disimilitud	52
II.6.2 Análisis de relaciones	53
II.6.3 Visualización de los datos	54
II.6.4 Modelos predictivos	54
II.7 Gestión y manejo de la minería de datos	56
II.7.1 Concepto de entrega del proyecto	56
II.7.2 Análisis del proyecto	56
II.7.3 Reunión del equipo para el proyecto	56
II.7.4 Metodología	57
II.7.5 Calendario del proyecto	57
II.7.6 Costo del proyecto	57
II.7.7 Entrega del proyecto	57
II.8 La minería de datos en la industria – Aplicaciones	58
II.8.1 Minería de datos en los servicios financieros	58

II.8.2 Minería de datos en las aseguradoras	59
II.9 La minería de datos y la inteligencia de negocios	60
III. TOMA DE DECISIONES	65
III. 1 Las decisiones y los encargados de tomarlas	65
III.2 Los tipos y estilos de decisiones	69
III.3 El proceso y los sistemas de soporte de la toma de decisiones	72
III.4 La toma de decisiones dentro de las organizaciones	75
III.5 Herramientas y modelos matemáticos para la toma de decisiones	82
III.5.1 Herramientas deterministas	82
III.5.1.1 Análisis de decisión lineal	82
III.5.1.2 Tablas de decisión	82
III.5.1.3 Análisis costo-beneficio	82
III.5.2 Herramientas complejas	82
III.5.2.1 Simulación	82
III.5.2.2 Decisiones con atributos múltiples	83
III.5.2.3 Sistemas de apoyo a las decisiones	83
III.5.2.3.1 Reglas de probabilidad	83
III.5.3 Modelos predictivos	84
III.5.4 Modelos de reconocimiento y aprendizaje de patrones	84
III.5.5 Modelos de optimización	84
III.5.6 Modelos de manejo de proyectos	85
III.5.7 Modelos de análisis de riesgo	85
IV. LA INVESTIGACIÓN DE MERCADO Y SUS APLICACIONES	86
IV.1 Definición e importancia de la investigación de mercados	86
IV.2 El proceso de investigación de mercados	88
IV.3 El Database Marketing	90
IV.4 La mercadotecnia de relaciones	93
IV.5 Inteligencia de mercadotecnia	97
V. CASO PRÁCTICO: IMPLEMENTACIÓN DE LA INTELIGENCIA DE NEGOCIOS	103
V.1 Antecedentes históricos y operativos	103
V.2 Análisis FODA y análisis IEM	107

V.3 Análisis de estrategias y propuestas de mejora utilizando inteligencia de negocios	110
V.3.1 Campañas de uso	110
V.3.1.1 Acuerdo sobre el propósito de la investigación	112
V.3.1.2 Objetivos de la investigación	113
V.3.1.3 Diseño de la investigación	113
V.3.1.4 Preparación y análisis de los datos	115
V.3.1.5 Resultados de la investigación y toma de decisiones	118
CONCLUSIONES	120
BIBLIOGRAFÍA	122

INTRODUCCIÓN

Desde hace ya varios años, el ámbito de los negocios se ha caracterizado por enfrentar dos principales retos: la fuerte competencia y el cambio rápido y constante. El Internet, se ha convertido en un arma muy poderosa para las empresas, pero puede también convertirse en un gran enemigo si no se tienen los conocimientos necesarios para explotarlo. El comercio electrónico y la necesidad de conocer los patrones de uso o compra de los clientes para poder crear estrategias enfocadas al cliente o a las relaciones comerciales, son necesidades imperantes en la actividad empresarial de nuestros días. Para responder de manera óptima a estos requerimientos las empresas buscan, de manera constante, respuestas y conocimiento, al interior y al exterior de ellas. Al mismo tiempo, acumulan grandes cantidades de datos sobre sus clientes y transacciones, que si no son utilizados de manera adecuada, pueden convertirse en repositorios de información sin ningún valor agregado.

Encontrar técnicas y herramientas que ayuden a la correcta utilización de los datos para después convertirlos en conocimiento que lleven a elaborar estrategias de mercadotecnia, relación con el cliente o negocios, debe ser una tarea primordial para todas las empresas que quieran tener o conservar una fuerte presencia en el mercado, ya que, de lo contrario, el cambio vertiginoso de éste, puede dejarlas en el rezago.

Existen varios factores que ayudan a detectar si una empresa determinada necesita un proyecto de Inteligencia de Negocios que apoye su evolución y sobre todo asegure resultados, minimizando las pérdidas resultantes de decisiones tomadas de acuerdo sólo a la experiencia o conocimiento de los empleados y generando nuevos clientes, al mismo tiempo que se conservan los ya existentes.

Una correcta metodología de Inteligencia de Negocios, implica saber utilizar modelos matemáticos y de análisis de datos para generar información y conocimiento útiles, y es básica para el correcto funcionamiento de las empresas. Usando esta metodología, se deben enfocar esfuerzos importantes en los procesos de Minería de Datos, para utilizar de la mejor manera los datos almacenados por las empresas y obtener información primordial y útil sobre los hábitos y preferencias de los clientes, orientados a enfocar los esfuerzos de todas las áreas de la empresa.

La Toma de Decisiones de manera rápida y correcta, puede ser el parte aguas entre las pérdidas y las ganancias económicas de una compañía y, finalmente, la Investigación de Mercados, es uno de los canales más importantes para dirigir las acciones hacia el elemento central de todos estos esfuerzos: el cliente.

Este trabajo describe primeramente qué es la Inteligencia de Negocios, cuáles son sus orígenes, su estructura y los tipos de análisis que se requieren para llevarla a cabo.

Como parte central de esta estructura, se desarrolla posteriormente la Minería de Datos, su funcionalidad, los tipos de datos y repositorios con los que se pueden trabajar, así como el proceso que conlleva a la utilización de las matemáticas para el posterior análisis de los datos que se obtengan.

Tomando como base estos dos temas, se aborda la Toma de Decisiones desde el punto de vista organizacional; quiénes son los encargados de tomarlas; los criterios, tipos y estilos de decisiones que se pueden realizar; cuáles deben ser los pasos para llegar a una correcta toma de las mismas y qué modelos matemáticos pueden ayudar en estas tareas.

Posteriormente, se elige el área de mercadotecnia de las empresas para puntualizar un área específica en la que se pueda utilizar la metodología anterior; la Investigación de Mercados y sus Aplicaciones como pieza fundamental de un proceso de trabajo enfocado al cliente, cuál es su importancia actual, el proceso que debe seguir dicha investigación, la combinación de la mercadotecnia con la minería de datos conocida como *Database Marketing* y más específicamente, la inteligencia de mercadotecnia en la que, nuevamente, se utilizan las matemáticas para definir estrategias dentro de esta área.

Una vez que se conoce la estructura anterior, se lleva a cabo un caso práctico. Se elige una empresa mexicana, con más de treinta años en el mercado, para realizar un análisis de la situación actual de la misma frente a la operación basada en la inteligencia de negocios, se detecta que hay varias áreas de oportunidad con respecto a esta metodología y se demuestra que, estructurando determinados esfuerzos y campañas con más y mejores análisis, se pueden evitar pérdidas y aumentar el grado de satisfacción de los clientes.

Si bien la implementación de un proyecto de esta naturaleza en una empresa puede generar gastos en el corto plazo, ya que se necesita contar con especialistas que analicen la información y con software especializado para el almacenamiento y explotación de los datos, los beneficios hacia el futuro serán mucho mayores ya que no se lanzarán esfuerzos ni campañas que no estén basados en información real, en análisis profundos y, sobre todo, no se tomarán decisiones sin fundamento alguno.

Seguramente, un proyecto de esta naturaleza generará una resistencia al cambio pero es importante sensibilizar a los colaboradores de la empresa, en que este tipo de proyectos son necesarios para la evolución de las operaciones, para evitar el trabajo que no genere un valor agregado y para optimizar los procesos conforme a los requerimientos del mercado.

I. INTELIGENCIA DE NEGOCIOS

La Inteligencia de Negocios se puede definir como el cúmulo de modelos matemáticos y metodologías de análisis que explotan los datos disponibles para generar información y conocimiento útiles para los complejos procesos de toma de decisiones.^[1]

La mayor parte de los colaboradores de una empresa, toman sus decisiones basándose en su experiencia, conocimiento y dominio de las aplicaciones así como la información disponible, que no son más que metodologías intuitivas y de fácil acceso; pero este tipo de decisiones no son óptimas en un ambiente económico que se está modificando constantemente. Es por ello que hoy en día los procesos de toma de decisiones en las empresas tienen que ser más complejos y dinámicos para poder enfrentar dichos cambios a través de un acercamiento intuitivo.

Los siguientes son ejemplos de dos procesos de toma de decisiones altamente complejos en condiciones de constante cambio:

- Retención de clientes en la industria de los teléfonos celulares. El gerente de mercadotecnia de una compañía de teléfonos celulares observa que una gran cantidad de clientes están dando de baja el servicio, debilitando a su empresa frente a sus competidores. Como puede imaginarse, la baja lealtad de los clientes, conocida como deserción es un factor crítico para muchas compañías que operan en la industria de servicios. Supongamos que el gerente de mercadotecnia puede confiar en un presupuesto adecuado para lograr obtener una campaña de retención de clientes de 2000 individuos de una base total de 2 millones de clientes. De aquí que la pregunta natural sea: ¿cómo se deben elegir dichos clientes para optimizar la eficiencia de la campaña? Dicho de otro modo, ¿cómo estimar la probabilidad de que cada cliente dé de baja el servicio para que entonces el blanco de la campaña sea de los mejores clientes para así reducir el nivel de deserción y maximizar la retención? Para conocer estas probabilidades, el blanco de clientes puede ser elegido como las 2000 personas que tienen el más alto nivel de deserción entre todos los clientes que representan un alto valor para el negocio. Sin el apoyo de modelos matemáticos avanzados y las técnicas de minería de datos sería muy complicado obtener una estimación confiable de la probabilidad de deserción para poder determinar los clientes potenciales de una determinada campaña de mercadotecnia.
- Planeación logística. El gerente logístico de una empresa manufacturera desea desarrollar un plan logístico de producción a mediano plazo. Este es un proceso de toma de decisiones de alta complejidad que incluye, entre otras decisiones, la asignación de la demanda originada desde diferentes áreas de mercado hacia los sitios de producción, la planeación de la producción en las plantas y la distribución de los productos terminados a las áreas de mercado. En una empresa manufacturera típica,

1. *Business Intelligence: Data Mining an Optimization for Decision Making/C Vercelis/Wiley and Sons Ltd/2009*

decenas de sitios de producción, cientos de proveedores y miles de productos terminados y componentes dentro de un marco de tiempo de un año, dividido en semanas. La magnitud y complejidad del problema sugiere que se requieren modelos de optimización avanzados para establecer el mejor plan logístico.

El propósito principal de los sistemas de Inteligencia de Negocios es proveer herramientas y metodologías a los trabajadores clave que les permitan tomar decisiones efectivas (planes de acción que permitan alcanzar los objetivos de una manera más efectiva) y eficientes (con la habilidad para reaccionar rápidamente frente a los competidores y a las nuevas condiciones de mercado tan cambiantes en nuestros días)

Si los encargados de la toma de decisiones pueden confiar en los sistemas de Inteligencia de Negocios facilitando su actividad, se puede esperar que el proceso de toma de decisiones global sea altamente mejorado, y por tanto, se puede concluir que la mayor ventaja derivada de la adopción de un sistema de Inteligencia de Negocios se encuentra en el aumento de la **efectividad** en el proceso de toma de decisiones.

I.1 BASES DE DATOS

Los datos que se originan, por una parte de transacciones internas de naturaleza administrativa, logística y comercial así como de fuentes externas, se acumulan dentro de los sistemas de información de organizaciones públicas y privadas. Las distintas fuentes contienen datos de calidad variable y usan representaciones inconsistentes, códigos y formatos que tienen que ser conciliados. Tales problemas de integración, limpieza y estandarización de los datos para preparar las tareas de Inteligencia de Negocios pueden ser un reto significativo. La carga de datos eficiente es imperativa para la Inteligencia de Negocios. Más aún, las tareas de Inteligencia de Negocios normalmente tienen que ser llevadas de manera progresiva conforme la llegada de nuevos datos, por ejemplo, las ventas del último mes. Esto genera una carga eficiente que actualiza las capacidades necesarias para poner en marcha las tareas de Inteligencia de Negocios además de existir una necesidad de soportar dichas tareas en tiempo real, esto es, la toma de decisiones de negocio basadas en los datos operativos per se.

Los datos sobre los que las tareas de inteligencia de negocios se llevan a cabo, están cargados típicamente en un repositorio llamado base de datos, el cual es manejado por uno o más administradores. Una elección muy popular de mecanismos para almacenar y programar los almacenes de datos son los sistemas de manejo de bases de datos relacionales (RDBMS por sus siglas en inglés). En los últimos 20 años, varias estructuras, optimizaciones y técnicas de procesamiento de datos se han desarrollado primordialmente para ejecutar una serie de consultas (*queries*) complejas en el lenguaje SQL sobre grandes cantidades de datos – un requerimiento clave para la Inteligencia de Negocios. Un ejemplo de dichas consultas en SQL sería: encuentra clientes que han solicitado una orden durante el último cuatrimestre y cuya cantidad exceda el volumen de orden promedio al menos en 50%.

Como principio básico, un programa sólido de Inteligencia de Negocios debe comenzar por un programa de calidad de datos formal posicionado en la base de cualquier esfuerzo. Las organizaciones que cuentan con programas formales de calidad de datos que soportan sus sistemas fuentes, tienen mayor probabilidad de tener éxito al establecer una estructura sólida de Inteligencia de Negocios. Este tipo de organizaciones entiende la importancia de asegurarse que la información sea confiable, precisa y que fluya a través de sus sistemas centrales. Las instituciones financieras sobresalen en esta área ya que la precisión de sus datos es lo que los mantiene en el negocio. Bases de clientes, ganancias, así como resúmenes de comportamiento de la cartera tienen que ser correctos ya que no se tiene una segunda oportunidad para tener errores y estas compañías saben que su credibilidad es vital para mantener las relaciones con sus clientes, así como su rentabilidad.

Tradicionalmente las bases de datos soportan funciones de análisis, que responden a dudas del negocio como las siguientes:

¿Cómo están las ventas del negocio? ¿Qué producto fue el mejor vendido en el último cuatrimestre?

¿Qué regiones tendrán el mejor y peor volumen de ventas respectivamente durante este año?

Estas son típicas dudas de negocio analíticas que normalmente son solicitadas por las áreas gerenciales quienes se inclinan por ver la tendencia del negocio desde un nivel alto y van profundizando en caso de ser necesario. Las herramientas de Inteligencia de Negocios en este ámbito proveen soporte al personal que maneja su negocio desde una perspectiva total y ayudan a señalar las áreas específicas que necesitan mejoras. El requerimiento de información es más acumulado y con más años de historia.

Por otra parte, existe una tendencia más reciente a usar las bases de datos y la tecnología de Inteligencia de Negocios para ayudar a mejorar la eficiencia en la operación para personas “normales” como los agentes de ventas, el representante de servicio al cliente o incluso a los propios clientes. Los requerimientos de datos son más oportunos e integrados. Estos usuarios buscan por el detalle de información integrada sobre personas específicas como un cliente al que se le hizo una llamada. Normalmente intentan contestar preguntas como:

¿Cuáles son las transacciones más recientes?

¿Cuál es el seguimiento actual a mis quejas? ¿Alguna de ellas fue rechazada?

¿Basados en la historia médica del cliente, cuál es el mejor plan de seguro que se adapta a sus necesidades? ¿Alguna recomendación?

Los tipos de usuarios arriba mencionados, no están interesados en el historial de transacciones a un nivel macro sino en transacciones más recientes y detalladas. El personal quisiera que las bases de datos pudieran proveer de dicha información de manera completa e integrada para ayudarlos a tomar decisiones tácticas. El servicio al cliente utilizaría esta información para sugerir o recomendar los siguientes pasos a los clientes. Estos servicios extras añaden excelencia en el servicio y mejoran las tasas de satisfacción del cliente. Comparadas con las necesidades desde el punto de vista analítico, los datos que se necesitan aquí son detallados, integrados y oportunos.

Normalmente el acceso a las bases de datos se basa en el uso de la misma y está dirigido a cierto grupo de gente en el negocio que tiene funciones específicas. Es muy difícil diseñar una sola base de datos que se acomode a las necesidades de estos dos diferentes tipos de usuarios. El reto es adaptar la arquitectura de la base de datos para ellos sin realizar grandes cambios en la misma.

Actualmente los negocios están haciendo frente a algunos temas que tienen que ver con requerimientos de información. La mayoría de las compañías tienen muchos sistemas de información distribuidos y los datos están dispersos por doquier. Es muy difícil para las personas ver todas las transacciones de datos con respecto a sus clientes de una vez, es decir, en una pantalla. El almacenamiento de datos parece ser el único lugar donde se puede tener la información de una forma tan detallada, depurada, integrada y completa. El cómo encaminar la actual inversión en almacenamiento de datos para ayudar al negocio a tomar decisiones inteligentes y objetivas y mejorar la eficiencia tanto de las necesidades analíticas como operacionales es otro reto para la gente de los departamentos de sistemas (IT)

La actividad de proveer soporte a los usuarios clave a través de la integración de los procesos de toma de decisiones y las tecnologías de información es conocida usualmente como gestión del conocimiento.

El propósito principal de la Inteligencia de Negocios así como de la gestión del conocimiento es desarrollar ambientes para proveer soporte a los usuarios clave en un proceso de toma de decisiones y actividades complejas para la resolución de problemas.

La Inteligencia de Negocios ya no se está utilizando sólo para crear reportes y análisis tácticos y estratégicos sino también para dirigir y optimizar los procesos de negocio en el día a día. La Inteligencia de Negocios ya no es un “nice to have” sino algo esencial en el éxito de los negocios.

I.2 MODELOS MATEMÁTICOS

Hoy en día es muy difícil encontrar empresas exitosas que no rijan sus negocios por la tecnología de Inteligencia de Negocios en los que recolectan datos más finos y precisos, cuyo volumen, por lo consiguiente, es mayor. Los negocios están utilizando sus datos agresivamente experimentando con técnicas de análisis de datos más sofisticadas que dirijan la toma de decisiones y arrojen nuevas funcionalidades como ofertas y servicios personalizados a los clientes a través de la aplicación de modelos matemáticos y algoritmos. En algunos casos, esta actividad se reduce al cálculo de totales y porcentajes, gráficamente representados por simples histogramas, mientras que análisis más elaborados requieren desarrollar modelos de optimización avanzada y aprendizaje.

En términos generales, la adopción de un sistema de Inteligencia de Negocios tiende a promover un acercamiento científico y racional en la gestión de las empresas. Este acercamiento racional, típico de un análisis de Inteligencia de Negocios, puede ser resumido sistemáticamente en las siguientes características:

- Primero, se identifican los objetivos del análisis y se definen los indicadores de comportamiento que se usarán para evaluar las distintas alternativas.
- Se desarrollan modelos matemáticos utilizando las relaciones entre las variables de control del sistema, parámetros y métricas de evaluación.
- Finalmente, los análisis *y si* son desarrollados para evaluar los efectos en el comportamiento determinado por los cambios en las variables de control y en los parámetros.

Entonces, la utilización de modelos matemáticos y abstractos en la Inteligencia de Negocios es muy importante ya que fuerzan a las personas encargadas de la toma de decisiones a enfocarse en los puntos clave de los objetivos analizados y así entender el fenómeno que se está investigando; el conocimiento se puede transferir a otros individuos dentro de la organización preservando la calidad del conocimiento y finalmente es tan general y flexible que en la mayoría de los casos puede ser aplicado a otras situaciones para resolver problemas similares.

I.3 ARQUITECTURA DE LA INTELIGENCIA DE NEGOCIOS

La figura I.1 describe los principales componentes de un sistema de inteligencia de negocios:



Fig. I.1 Componentes de un sistema de inteligencia de negocios

I.3.1 Fuentes de Datos: Como primer plano se deben reunir e integrar los datos almacenados en fuentes primarias y secundarias heterogéneas en cuanto su origen y su tipo. Las fuentes de datos consisten en su mayoría en datos pertenecientes a sistemas operativos pero también pueden incluir documentos no estructurados como correos electrónicos y datos recibidos de proveedores externos.

I.3.2 Almacenes de Datos (Data warehouses y data marts): Utilizando herramientas de extracción e información conocidas como herramientas ETL (extraer, transformar y cargar por sus siglas en inglés), los datos originados de las diferentes fuentes son almacenados en bases de datos que pretenden soportar los análisis de Inteligencia de Negocios.

I.3.2.1 Data warehouse: Como su nombre lo indica, un almacén de datos es el principal repositorio de los datos disponibles para desarrollar arquitecturas de inteligencia de negocios y sistemas de apoyo a las decisiones. Existen tres categorías principales de datos que alimentan un almacén de datos: datos internos, datos externos y datos personales.

I.3.2.1.1 Datos internos: La mayor parte de estos datos se encuentran en las bases de datos transaccionales u operacionales que son el sistema medular de información de una empresa. Normalmente están reunidos a través de aplicaciones que soportan la operación de rutina de una compañía referente a la administración, contabilidad, producción y logística. El conjunto de estas aplicaciones de índole transaccional se conocen como ERP (Enterprise Resource Planning). Estos datos vienen normalmente de diferentes componentes de los sistemas de información:

Sistemas back-office: Registros transaccionales de órdenes, facturas, inventarios y datos logísticos y de producción.

Sistemas front-office: Datos originados de las actividades de un call center, servicio a clientes, ejecución de campañas de mercadotecnia.

Sistemas web-based: Reúnen transacciones de venta a través de comercio electrónico, visitas a sitios web, datos disponibles en formatos que son llenados por clientes y prospectos.

I.3.2.1.2 Datos externos: Se utilizan para enriquecer la información almacenada en las bases de datos internas. Algunas agencias, por ejemplo, reúnen y hacen disponibles los datos relativos a las ventas, al mercado y a predicciones de tendencia futuras para algunos negocios específicos así como indicadores económicos y financieros. Otro ejemplo es la información que proveen los sistemas de información geográfica que organizan, almacenan y presentan datos de acuerdo a una división territorial.

I.3.2.1.3 Datos personales: En la mayoría de los casos, los encargados de la toma de decisiones, también confían en información y habilidades personales almacenadas en hojas de trabajos o bases de datos locales resguardadas en sus propias computadoras. Lo importante al utilizar este tipo de información es cómo se combina e integra con los datos internos y externos para tomar decisiones pertinentes.

I.3.2.2 Data Marts: Los data marts son subconjuntos de datos que reúnen toda la información requerida por un departamento específico en la compañía como pueden ser las áreas de mercadotecnia o logística, con el propósito de desarrollar análisis de inteligencia de negocios y ejecutar aplicaciones de apoyo a las decisiones específicas a dichas funciones. Por lo tanto, un data mart puede ser considerado como un almacén de datos más funcional o referente a un departamento, de un menor tamaño y tipo más específico que un almacén de datos de la compañía en general. Los data marts contienen un subconjunto de datos almacenados en el almacén de datos de la empresa y que se integran y cruzan con los datos de otros departamentos para obtener informaciones específicas.

Los almacenes de datos deben de tener las siguientes características:

- Orientados a entidades (*entity oriented*): Los datos contenidos en un almacén de datos tienen que ver principalmente con las entidades que poseen intereses en el análisis como los productos, clientes, órdenes y ventas.
- Integrados: Los datos tienen que ir siendo integrados y homogeneizados conforme se van cargando en el almacén de datos.
- Variable en el tiempo: Todos los datos que se cargan al almacén de datos, deben ser etiquetados conforme al periodo de tiempo al que hacen referencia ya que la medida de tiempo en cualquier almacén de datos es un elemento crítico que juega un rol muy importante.
- Persistentes: Una vez cargados en el almacén de datos, estos no deben ser modificados y se deben mantener y resguardar permanentemente.
- Consolidados: Para reducir espacio de almacenamiento a través de los años ya que normalmente, gran parte de los datos almacenados se obtienen como cargas parciales pertenecientes a los sistemas operativos que las originaron.
- No-normalizados (*desnormalized*): Los datos almacenados en un almacén de datos no están estructurados de una manera normal sino que pueden permitir redundancias para llevar a tiempos de respuesta menores en queries complejos.

1.3.3 Exploración de Datos: Herramientas para desarrollar un análisis de Inteligencia de Negocios pasivo (ya que el encargado de la toma de decisiones debe generar hipótesis previas o definir criterios para la extracción de los datos) que consiste en utilizar consultas y sistemas de reporte así como métodos estadísticos. El propósito principal del análisis exploratorio de datos es resaltar las características relevantes de cada atributo contenido en el conjunto de datos, utilizando métodos gráficos y calculando estadísticas resumen así como identificar la intensidad de las relaciones subyacentes primordiales entre estos atributos. El análisis exploratorio de los datos incluye tres fases principales:

1.3.3.1 Análisis univariado: Análisis en que se investigan las propiedades de cada uno de los atributos del conjunto considerándose cada uno de ellos independientes de otras variables del conjunto de datos. Uno de los intereses es evaluar la tendencia de los valores de un atributo dado para organizarlos alrededor de un valor central específico (ubicación), medir la propensión de la variable a asumir un rango de valores más o menos amplio (dispersión) y extraer información de la distribución de probabilidad subyacente.

El análisis univariado tiene varios objetivos. Por una parte, algunos modelos hacen hipótesis estadísticas específicas con respecto a la distribución de las variables que son examinadas y por ello es necesario

verificar la validez de cada una de las premisas antes de proceder con las investigaciones subsecuentes. Por otro lado, el análisis univariado grafica intuitivamente conclusiones que tienen que ver con el contenido de información que cada atributo puede proveer.

El análisis univariado puede ser tan sencillo como una gráfica de barras hasta medidas de ubicación o tendencia central, dispersión y ubicación relativa:

- Ubicación: Dentro de las principales estadísticas de ubicación que también son llamadas medidas de tendencia central tenemos:
 - Media
 - Mediana
 - Moda
 - Rango medio
 - Media geométrica
- Dispersión: Para representar el nivel de variabilidad expresado en las observaciones con respecto a los valores centrales tenemos:
 - Rango
 - Desviación media absoluta
 - Varianza
 - Coeficiente de variación
- Ubicación relativa: Cuantiles, cuartiles, percentiles.

1.3.3.2 Análisis bi variado: Análisis en que se consideran parejas de atributos para explorar la la relación existente entre ellos. Al igual que en análisis univariado, podemos recurrir tanto al análisis gráfico como a algunos indicadores.

- Análisis Gráfico: Permiten la visualización de la relación entre dos atributos.
 - Gráficos *Scatter* (de dispersión): Es un plano cartesiano bi-dimensional que se obtiene colocando el atributo a_j en el eje horizontal y el segundo atributo a_k en el eje vertical. Los puntos graficados resultantes corresponden a las parejas de valores (x_{ij}, x_{ik}) , $i \in M$, del conjunto de datos D .

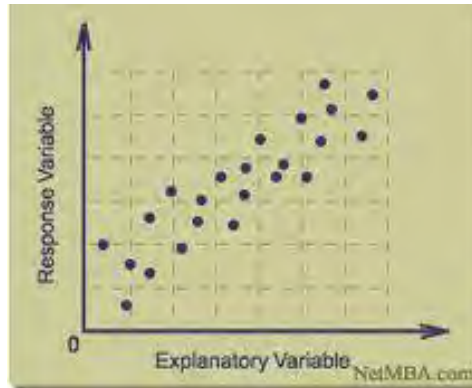


Fig. I.2 Gráfico Scatter

- Gráficos *Loess*: Están basados en los gráficos *scatter* y pueden ser aplicados alternativamente a pares de atributos numéricos. Comenzando con un gráfico *scatter*, es posible añadir una curva de tendencia para expresar la relación funcional entre el atributo a_k y el atributo a_j . La curva de tendencia se puede obtener utilizando técnicas de regresión, lo que explica el término *loess*: *local regression* (regresión local en inglés)

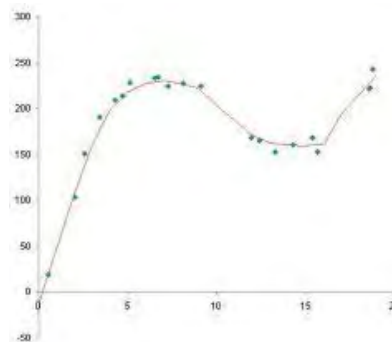


Fig. I.3 Gráfico Loess

- Curvas de nivel: Gráficos con un mayor desarrollo que los gráficos *scatter* y que pueden ser usados solamente para atributos numéricos. Resaltan el valor de un tercer atributo numérico a_z , mientras que los atributos a_j y a_k colocados en los ejes del gráfico varían. Conectando entre sí los puntos del gráfico que comparten el valor del tercer atributo, posiblemente utilizando algún tipo de interpolación numérica, se obtienen líneas curvas que representan el lugar geométrico de los puntos para los que el atributo a_z asume un valor dado. Un ejemplo de

curvas de nivel son las *líneas de contorno* que se obtienen resaltando en un mapa geográfico los puntos que comparten la misma altitud, lo que correspondería en este caso al atributo a_z .

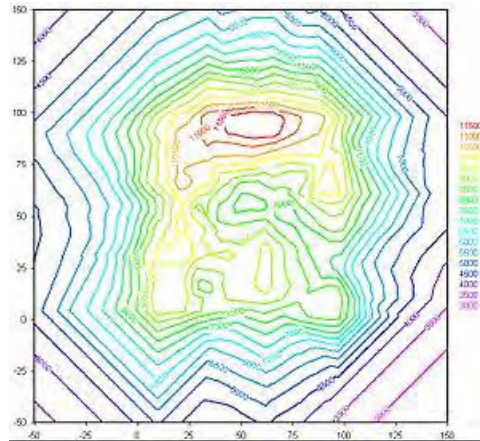


Fig. I.4 Curvas de nivel

- Gráficos de Cuantiles y Cuartiles: Conocidos con *QQplots* por sus siglas en inglés, se utilizan para comparar las distribuciones de un mismo atributo para dos diferentes características de la población o para muestras extraídas de dos diferentes poblaciones. El análisis es aplicable a atributos numéricos y se lleva a cabo al comparar los cuantiles de dos series de observaciones.

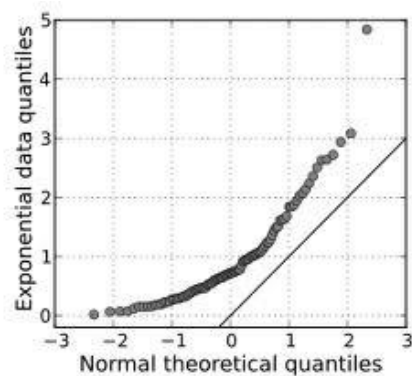


Fig. I.5 Gráfico de Cuantiles y Cuartiles

- Series de tiempo: Existen algunas situaciones en las que el atributo a_k representa el valor numérico de una cantidad medible, por ejemplo las ventas de un producto, mientras

que el atributo a_j indica la secuencia de periodos de tiempo en el que los valores de a_k fueron recolectados. Tales periodos corresponden a medidas naturales de tiempo como horas, días, meses, cuatrimestres y años. Son, por lo tanto variables discretas y se colocan en intervalos regulares a través del eje horizontal.

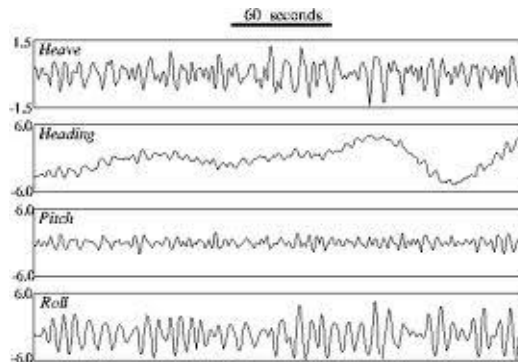


Fig. I.6 Series de Tiempo

- Indicadores resumen: Expresan la naturaleza y la intensidad de la relación entre atributos numéricos. Entre estos indicadores tenemos:
 - Covarianza
 - Correlación

1.3.3.3 Análisis multivariado: Análisis en que se investigan las relaciones existentes entre un subconjunto de atributos. También se cuentan con herramientas gráficas y medidas de correlación.

- Análisis Gráfico: Todos estos métodos funcional exclusivamente para atributos numéricos:
 - Matriz de gráficos scatter: Son matrices de gráficos evaluadas para cada par de variables numéricas. De esta manera, es posible visualizar la naturaleza y la intensidad de las relaciones por parejas en un solo gráfico.

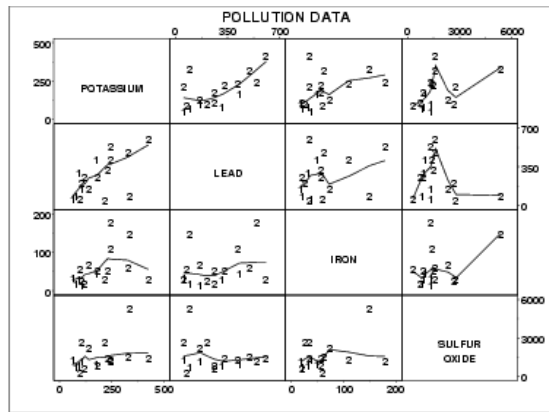


Fig. I.7 Matriz de gráficos scatter

- Gráficos de estrella: Muestran de una manera intuitiva las diferencias entre valores de los atributos para los registros de un conjunto de datos. Para que sean efectivos, deben ser aplicados a un número limitado de observaciones, por ejemplo no más de una docena, y la comparación debe estar basada en un número pequeño de atributos. El concepto básico involucra relacionar cada registro con un ícono en forma de estrella de cuyo centro parten tantos rayos como el número de atributos. La longitud de cada rayo es igual al valor del atributo correspondiente, normalizado para caer en el intervalo [0,1] para que arroje una representación consistente de los varios atributos.

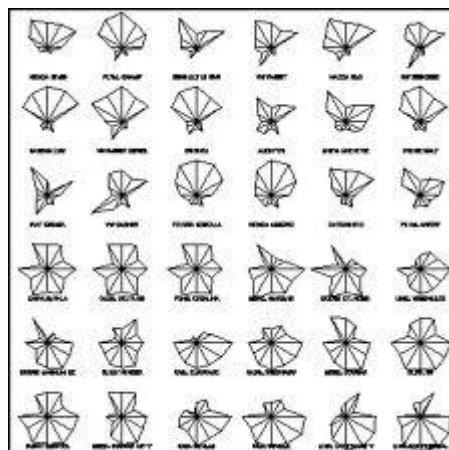


Fig. I.8 Gráficos de estrella

- Medidas de correlación: Para el análisis multivariado de atributos numéricos, las matrices de covarianza y correlación se calculan entre todos los pares de atributos.

Antes de proceder con la exploración de los datos, es necesaria la preparación de los mismos con el fin de contar con un conjunto de datos de alta calidad para usos subsecuentes en la inteligencia de negocios y el análisis de minería de datos. La calidad de estos datos es una preocupación constante de aquellas personas responsables del diseño y la actualización de los almacenes de datos. Los principales factores que pueden afectar la validez, integridad y calidad de los datos son:

- **Precisión:** Para que la información sea útil en análisis subsecuentes
- **Datos completos:** Un gran número de datos faltantes pueden comprometer la precisión con que el negocio se maneja.
- **Consistencia:** En forma y contenido dado que los datos son tomados de distintas fuentes durante el proceso de integración de la información.
- **Tiempo:** Los datos deben actualizarse frecuentemente basados en los objetivos del análisis. Se recomiendan actualizaciones diarias o máximo semanales.
- **Redundancia:** La repetición de datos lleva a una pérdida de espacio en la memoria y a la inconsistencia en la información.
- **Relevancia:** Los datos deben ser relevantes de acuerdo a las necesidades del sistema de inteligencia de negocios para añadir valor real a los análisis que se llevarán a cabo.
- **Interpretación:** El significado de los datos debe ser bien entendido y correctamente interpretado por los analistas.
- **Accesibilidad:** Los datos deben ser fácilmente accesibles para los analistas y las aplicaciones de soporte en la toma de decisiones.

1.3.4 Minería de Datos: Metodologías activas de Inteligencia de Negocios cuyo propósito es la extracción de la información y el conocimiento de los datos. La minería de datos incluye modelos de reconocimiento de patrones, aprendizaje en cuanto a las máquinas y técnicas de minería de datos y constituye un proceso iterativo enfocado al análisis de grandes bases de datos con el propósito de extraer información y conocimiento que pueda proveer de conocimiento preciso y potencialmente útil para los trabajadores comprometidos en la toma de decisiones y la resolución de problemas. El capítulo II describirá y detallará las técnicas, conceptos y aplicaciones de la minería de datos.

1.3.5 Optimización: Determinar la mejor solución dentro de un cúmulo de alternativas de acción, que muchas veces es muy extenso y en algunas ocasiones, de tamaño infinito. Modelos de optimización aplicados a la mercadotecnia serán descritos en el capítulo III y IV.

1.3.6 Decisiones: Finalmente, la parte más alta de la pirámide corresponde a la elección y adopción de una decisión específica y que, de alguna manera, representa la conclusión natural de un proceso de toma de decisiones.

I.4 CICLO DEL ANÁLISIS DE INTELIGENCIA DE NEGOCIOS

Es posible identificar un ciclo ideal que caracteriza la evolución de un análisis típico de Inteligencia de Negocios:

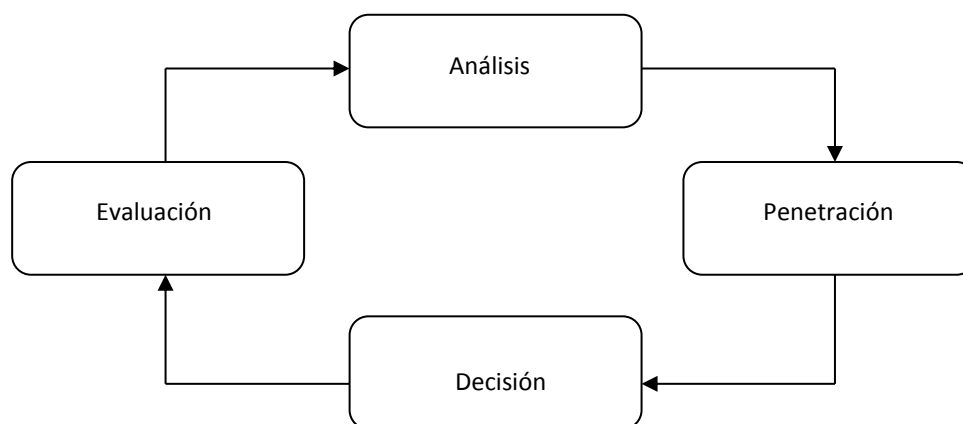


Fig. I.9 Ciclo de análisis de Inteligencia de Negocios

I.4.1 Análisis: Durante la fase de análisis, es necesario reconocer y detectar el problema. Los encargados de la toma de decisiones deben crear una representación mental del fenómeno que se analizará, identificando los factores críticos que se perciban como relevantes. Las metodologías de Inteligencia de Negocios ayudan en este punto, permitiendo a los encargados de la toma de decisiones desarrollar patrones de investigación.

I.4.2 Penetración: La segunda fase permite a los encargados de la toma de decisiones entender mejor y más profundamente el problema. La información obtenida dentro de la fase de análisis se transforma en conocimiento durante la fase de penetración.

I.4.3 Decisión: Durante la tercera fase, el conocimiento obtenido como resultado en la fase de penetración se convierte en decisiones y subsecuentemente en acciones. La disponibilidad de metodologías de Inteligencia de Negocios permiten que las fases de análisis y penetración sean ejecutadas de manera más rápida para que entonces se puedan tomar decisiones más efectivas y eficientes que encajen en las prioridades de una determinada organización.

I.4.4 Evaluación: Finalmente, la cuarta fase del ciclo de Inteligencia de Negocios contempla la medición del comportamiento y la evaluación. Mediciones extensas deben ser ideadas para que no se limiten sólo a los aspectos financieros sino también deben tomar en cuenta los indicadores de comportamiento más importantes definidos por los diferentes departamentos en la empresa como son la redención de las campañas de mercadotecnia, tamaño de las muestras, eficiencias en los equipos de trabajo, análisis de tasas de cambio en morosidad (*roll rate*) y edad (*aging*) de la cartera, etc.

En resumen, un sistema de Inteligencia de Negocios puede desarrollarse siguiendo las siguientes etapas:

- **Análisis:** De la información almacenada en las bases de datos de la empresa y en la que los datos hayan sido homogeneizados y estandarizados con ayuda de las tecnologías de información, previos requerimientos de los encargados de la toma de decisiones, tomando en cuenta las necesidades de las diferentes áreas de la empresa: operativos y analíticos.
- **Diseño:** En el que se desarrolle el plan de trabajo como tal tomando en cuenta los tiempos de respuesta, costos y prioridades de la empresa en todos los niveles.
- **Planeación:** Es en esta etapa donde entra la adopción de los modelos matemáticos, se revisa qué datos se utilizarán para cada uno de ellos y se verificará la eficiencia de los algoritmos que ayudarán a la resolución del problema.
- **Implementación y control:** En esta etapa se lanza el sistema de Inteligencia de Negocios a toda la empresa con su respectivo control en todos los ámbitos para asegurarse de que los resultados esperados sean obtenidos y utilizados dentro de las mejores prácticas para la toma de decisiones.

I.5 SOFTWARE

Uno de los factores más críticos dentro del proceso de Inteligencia de Negocios es el software que puede ser una colección de tecnologías de apoyo a las decisiones para la empresa que encamina y facilita el conocimiento a los miembros de la misma: ejecutivos, gerentes y analistas para tomar decisiones rápidas y mejores. En los últimos años se ha visto un crecimiento exponencial tanto en el número de productos y servicios que se ofrecen y en la adopción de estas tecnologías en la industria. Este crecimiento ha sido impulsado por la baja de los costos de adquirir y almacenar grandes cantidades de información que nacen de fuentes como transacciones bancarias de clientes, servicios financieros para análisis de quejas y detección de fraudes, en telecomunicaciones para identificar razones para la deserción de clientes, así como comercio electrónico, email, blogs y calificaciones de productos varios.

Existen varias aplicaciones “front-end” muy populares a través de las cuales los usuarios desempeñan las tareas de Inteligencia de Negocios: hojas de cálculo, portales de búsqueda, aplicaciones de manejo de desempeño que permiten a los encargados de la toma de decisiones dar seguimiento a indicadores claves de desempeño del negocio utilizando tableros y herramientas que permitan a los usuarios formular queries ad hoc y modelos de minería de datos. Rápidamente, una visualización ad hoc de los datos, pueden permitir la exploración dinámica de patrones y ayudan a descubrir factores relevante para la inteligencia de negocios.

El apoyo a las plataformas de inteligencia de negocios en medio de grandes organizaciones no tiene que ver sólo con tecnología o la información que la tecnología proporciona. Encontrar el balance exacto entre el tipo de información requerida y el marco de la entrega requiere un acercamiento definido y metódico construido sobre principios sólidos gobernanza (eficacia, calidad y buena orientación de los resultados) que combinen las premisas de negocio con las tecnologías apropiadas que lo permitan.

I.6 INTELIGENCIA DE NEGOCIOS EN TIEMPO REAL

Hoy en día, los Data marts y las soluciones de inteligencia de negocios que operen en tiempo real son cruciales para las organizaciones para asegurar su permanencia y competitividad en su negocio. Entre los retos para implementar una solución de inteligencia de negocios en tiempo real se encuentran:

- La habilidad para proveer información analítica con un funcionamiento superior mientras se actualizan los data marts en tiempo real.
- La justificación costo-beneficio de la inteligencia de negocios en tiempo real
- El impacto en el comportamiento y funcionamiento de una o más aplicaciones fuente.

En términos absolutos, no existe, como tal, una solución de inteligencia de negocios en tiempo real. La definición de “tiempo real” es relativa a las necesidades de información críticas dentro de una organización: para algunas se puede definir con la información que se obtiene dentro de un minuto de tiempo mientras que para otras, se requiere una respuesta instantánea, en periodos que pueden ser tan pequeños como unas cuantas milésimas de segundo. Para cada uno de estos escenarios se debe incluir el tiempo que se toma en capturar los cambios en los sistemas fuente, poner a funcionar los procesos ETL y almacenar los reportes a los que se accesan frecuentemente. Al considerar el tiempo que se toma cada paso de la integración de los datos y la presentación de la solución nos podemos aproximar a un “tiempo real cercano”. Sin embargo, muchas de las soluciones de inteligencia de negocios se enfrentan con frecuencia con información que se necesita tener al minuto o que se puede tardar máximo algunos minutos.

La mayoría de las iniciativas de la inteligencia de negocios en tiempo real, comienza con el objetivo de entregar la información correcta en el tiempo correcto. Esto describe correctamente la importancia de los periodos de tiempo y corrección de la información, sin embargo una solución de inteligencia de negocios en tiempo real que arroje información al minuto a través de reportes y tableros puede no ser suficiente. Para un vicepresidente de ventas que con dificultad cuenta con el tiempo de ingresar a su laptop para acceder a la información cuando está visitando a clientes clave, esta información es tan buena como no disponible. Para usuarios así, la solución no se puede etiquetar como “en tiempo real” en el sentido más estricto al menos que también sea entregada en el sitio correcto donde sea fácilmente accesible.

La minería de datos y análisis ad-hoc satisfactorios son piezas claves del almacenamiento de datos. Ambos tipos de análisis tienen impacto sobre el funcionamiento de la base de datos al mismo tiempo. Para una solución en tiempo real, esto es un reto primordial ya que la solución tiene que trabajar tanto sobre la actualización de los datos como con la entrega de reportes en tiempo real. Si un usuario utiliza la base de datos con una consulta que necesite unir todas las dimensiones de las tablas de hechos mientras que los procesos ETL están trabajando sobre actualizaciones en tiempo real, el comportamiento de la base de datos se verá seriamente deteriorado. Es por esto que los límites de la minería de datos y las necesidades de los análisis ad-hoc tienen que definirse durante la fase de requerimientos.

Pero esto nos lleva a uno de los mitos acerca de las soluciones de inteligencia de negocios y es que sólo sirven para soportar necesidades de reportes de carácter analítico mientras que otras aplicaciones transaccionales arrojan información de carácter operacional. La realidad es que una solución de inteligencia de negocios debe satisfacer ambos tipos de necesidades de igual manera. La inteligencia de negocios en tiempo real no es la excepción. Tableros y alertas no están destinados sólo para los niveles gerenciales o directivos. Típicamente, en una organización, la mayoría de los reportes analíticos son generados e interpretados por los niveles medios de la empresa quienes son, al mismo tiempo, responsables de monitorear el comportamiento y eficiencia de las áreas de operaciones en tiempo real.

Con cambios en tiempo real que se reflejen en soluciones de inteligencia de negocios, toda una nueva plataforma se abre a los usuarios para que puedan acceder a la información de operaciones al minuto. Esto supera las soluciones de inteligencia de negocios tradicionales donde los usuarios tienen que esperar un día o más para tener la misma información que contiene la aplicación fuente.

I.7 EJEMPLOS DE TÉCNICAS PARA INTEGRAR LA INTELIGENCIA DE NEGOCIOS EN UNA EMPRESA

Es posible integrar la inteligencia de negocios en los procesos del negocio de diversas maneras, como por ejemplo:

- Integración de las aplicaciones analíticas con las operacionales utilizando un portal empresarial para el acceso y explotación de los usuarios internos y externos.
- Incrustar las aplicaciones analíticas en las operacionales durante el desarrollo de la aplicación.
- Introducir servicios web para integrar de una manera dinámica el procesamiento analítico con las aplicaciones internas y operacionales para ayudar al beneficio comercial de colaboración.

I.7.1 PORTALES EMPRESARIALES

Un portal empresarial provee a usuarios internos y externos de una sola interfaz, segura y en red que personaliza contenido integrado, es decir, que el portal integra aplicaciones (operacionales o analíticas, internas o externas), información (datos estructurados, reportes, gráficos, cubos, etc.) así como contenido interno o externo no estructurado (documentos, medios digitales, etc.) y herramientas de colaboración (chat vía web, reuniones virtuales, email, mensajería instantánea, etc.) en una sola interfaz de usuario.

Un portal empresarial no es lo mismo que una intranet. Una diferencia clave entre ambos es que la tecnología del portal ofrece personalización mientras que la intranet no lo ofrece. Cada usuario ve algo diferente dependiendo de su rol dentro de la organización. El portal empresarial es un producto de una tecnología específica que ofrece mucho más que una interfaz de usuario vía web. Los portales reproducen contenido de sistemas subyacentes en la pantalla como un conjunto de pequeños componentes modulares o *portlets* que hacen que parezca que todo está en un solo sistema. Estos componentes modulares son simplemente partes de la pantalla que dan al usuario el acceso a la funcionalidad de la aplicación, la inteligencia de negocios, las herramientas de colaboración, contenido no estructurado, etc.

También pueden existir varios portales en donde cada uno consiste en un conjunto de componentes modulares que arrojan información de diferentes sistemas. Estos diversos portales pueden ser diseñados para que no sólo estén personalizados de acuerdo a un rol específico sino a tareas específicas, es decir, los portlets en un portal pueden ser asociados a tareas particulares del proceso de negocio llevadas a cabo por una persona en un rol específico. Más aún, los portales son un dispositivo independiente y pueden hacer frente junto con los usuarios que ingresan a las aplicaciones de inteligencia de negocios en dispositivos móviles.

Los servicios web hacen más fácil la integración de la inteligencia de negocios en los productos que ofrecen portales así como la integración de los portales de inteligencia de negocios en los portales empresariales. Algunos proveedores de estas aplicaciones son SAS y MicroStrategy que cuentan con soporte de servicios web en sus productos, lo que permite publicar como servicios web las herramientas y objetos de inteligencia de negocios como son reportes, análisis de cubos, modelos de minería de datos, etc.

Esto nos lleva a concluir que la integración de las aplicaciones de inteligencia de negocios en portales es una manera rápida y económica de proporcionar una herramienta de inteligencia de negocios basada en roles, a un grupo masivo de usuarios. Además, la implementación de portales con tableros, indicadores y herramientas de colaboración es muy conveniente para los ejecutivos, gerentes y usuarios de alto nivel jerárquico que necesitan crear estrategias, evaluar y colaborar antes de tomar una decisión. Estos portales no deben ser desarrollados de una manera demasiado cerrada y enfocada para que puedan dar soporte a cualquier tipo de información del negocio, aplicación o servicio que utilice este tipo de tecnología.

1.7.2 INCRUSTAR LAS APLICACIONES ANALÍTICAS EN LAS OPERACIONALES

Esta actividad normalmente es enfrentada por la gente de desarrollo del área de sistemas en la empresa (IT) durante el desarrollo de la aplicación. El equipo de desarrollo puede elegir cómo hacer esto: explotar la funcionalidad de inteligencia de negocios al nivel de la aplicación, al nivel de los datos o ambos.

Para hacerlo al nivel de la aplicación se pueden utilizar las interfaces de programación de aplicaciones (APIs), en el que las aplicaciones operacionales pueden utilizar este acercamiento para “llamar” herramientas de inteligencia de negocio como son reportes, cubos, análisis automáticos, etc. El equipo de desarrollo puede simplemente codificar las APIs en las aplicaciones operacionales o alternar y utilizar las herramientas desarrolladas previamente para seleccionar la funcionalidad que necesiten de una biblioteca de componentes de inteligencia de negocios pre construidos que incluyen procesos de análisis en línea, minería de datos, navegación y visualización de los componentes.

Las ventajas en esta técnica incluyen:

- La integración estrecha de la funcionalidad de inteligencia de negocios en una aplicación operacional significa que el análisis es usado en el mismo contexto de negocios que los datos operacionales.
- Este acercamiento es particularmente diseñado para los consumidores de la información (clientes, equipo de operación directa) quienes, sin embargo, están atados a una interfaz de usuario específica.

- Al usuario se le presenta una interfaz consistente de usuario y operación: el usuario no tiene que salir de su aplicación operacional para acceder al análisis.
- Las funciones analíticas incrustadas en las bases de datos soportan la explotación de inteligencia de negocios dentro de las aplicaciones de la empresa.

Por otro lado, las desventajas pueden ser:

- Se requiere del equipo de sistemas (IT) para que desarrolle la integración
- Sólo las nuevas aplicaciones podrían tener ventajas en esta opción
- Sería difícil separar las funciones de análisis en el futuro si surgiera el requerimiento.

I.8 BARRERAS ORGANIZACIONALES A LA INTELIGENCIA DE NEGOCIOS

Las organizaciones están sintiendo la presión de un clima económico estancado, regulaciones gubernamentales cambiantes, requerimientos de nuevos clientes, temas de seguridad y servicios externos. En respuesta, se están aferrando a soluciones tecnológicas para detener su inminente pérdida de conocimiento en el negocio y sin embargo se siguen encontrando en el mismo sitio, sólo que con una nueva plataforma.

La razón por la que muchas aplicaciones de inteligencia de negocios fallan no se debe a la tecnología. En gran medida, estas aplicaciones fallan debido a fallas organizacionales, culturales y de infraestructura en las empresas. La tecnología sólo puede dirigirnos en cómo hacer las cosas, sin embargo, las causas de las fallas están en lo que hacemos y lo que no.

La mayor parte de estas barreras están basadas en el rechazo o miedo al cambio, por lo que es necesario conocer y atacar este miedo por medio de la inteligencia de negocios. Algunos de los cambios más temidos son:

- **Cambios en la jerarquía:** Las iniciativas de inteligencia de negocios normalmente sacuden diversas posiciones en la jerarquía organizacional con la implementación de nuevos procesos y tecnologías que puedan hacer que las personas encargadas de cierta tarea no tengan ya el mismo reconocimiento por su trabajo y por lo tanto no puedan permanecer en el puesto que tienen o no les sea tan fácil saltar al nivel inmediato superior.
- **Cambios en la cultura laboral:** En una empresa con una filosofía de inteligencia de negocios, todos los usuarios toman la responsabilidad de tomar las decisiones del negocio lo que puede causar temor entre la mayoría de los trabajadores. La inteligencia de negocios tiene una justificación de negocio y no sólo una justificación tecnológica. Un gerente de proyecto ya no es el único encargado de establecer los objetivos e indicarlo a los usuarios sino que los usuarios mismos tienen que aprender a establecer las metas de los proyectos y después negociarlas con el gerente. Los usuarios tendrán que elegir directamente y desde el momento de la implementación qué datos necesitarán para posteriormente llevar a cabo sus análisis respectivos y llevar a cabo la toma de las decisiones. De igual manera, al tener este control y responsabilidad con respecto a la información, las expectativas de los usuarios se modificarán lo que puede generar tensión en el ambiente de trabajo con actitudes como “Los demás no entregan la información a tiempo”, “Esperan que esta filosofía resuelva todos los problemas”, “Trabajamos contra reloj pero siempre piden más”.
- **Desplazamiento de autoridad:** Al implementar una filosofía de inteligencia de negocios así como una nueva tecnología o sistema basado en dicha filosofía, la autoridad de los proyectos pasa de manos del equipo de sistemas (IT) hacia el lado de negocios lo que provoca cierto temor a la pérdida de tareas de un lado y al aumento de responsabilidades por el otro.
- **Reclutamiento:** Al incluir a personal interno en la implementación se corre el riesgo de que algunas tareas claves dentro de la organización se descuiden si a los colaboradores se les pide que ocupen el 100% del tiempo en el proyecto. Por otro

lado, si los niveles superiores no comprenden la importancia de dedicar el 100% del tiempo, se corre el riesgo de sobre saturar de responsabilidades a los encargados de la implementación y si, además, no se otorgan los incentivos correctos y necesarios, se creará un ambiente de inconformidad que no ayudará al avance en la implementación. Si se contrata personal externo para la implementación habrá molestias en cuanto a lo que se les está pagando a los externos y también se corre el riesgo de que crezca el presupuesto destinado al proyecto.

II. MINERÍA DE DATOS (DATA MINING)

II.1 ¿QUÉ ES LA MINERÍA DE DATOS?

La minería de datos se refiere a la extracción de conocimiento a partir de grandes cantidades de datos, sobre los que se aplican métodos inteligentes con el fin de extraer patrones interesantes. Dichos datos pueden estar almacenados en bases de datos, almacenes de datos u otros repositorios de información.

El proceso de descubrimiento del conocimiento incluye la limpieza, integración, selección, transformación y minería de los datos para una posterior evaluación de patrones y presentación del conocimiento.

La minería de datos, es un campo interdisciplinario que nace a partir de áreas como los sistemas de bases de datos, almacenamiento de datos, estadística, visualización de datos, recuperación de información y programación avanzada. Otras áreas que han contribuido al desarrollo de la minería de datos son las redes neuronales, el reconocimiento de patrones, el análisis de datos espaciales, las bases de datos de imágenes, el procesamiento de señales y muchos otros campos de aplicación como son los negocios, la economía y la bioinformática.

Los patrones de datos pueden ser *minados* a partir de diferentes tipos de bases de datos como son las bases de datos relacionales, los almacenes de datos o *data warehouses* y las bases de datos transaccionales, las bases de relación de objetos y las orientadas a objetos respectivamente. Patrones de datos interesantes, también pueden ser extraídos de otros tipos de repositorios de información como las bases de datos espaciales, de relación de tiempo, de texto, multimedia y de la propia red global.

La mayor razón por la que la minería de datos ha atraído la atención en la industria de la información en los últimos años se debe a la amplia disponibilidad de grandes cantidades de datos en las empresas y a la inminente necesidad de convertir dichos datos en información y conocimiento útiles. La información y el conocimiento obtenidos pueden ser utilizados para aplicaciones que van desde la gestión de un negocio, el control de producción y el análisis de mercado hasta el diseño en ingeniería y la exploración científica.

La minería de datos puede ser vista como resultado de una evolución natural de la tecnología de la información como se muestra en la figura II.1:

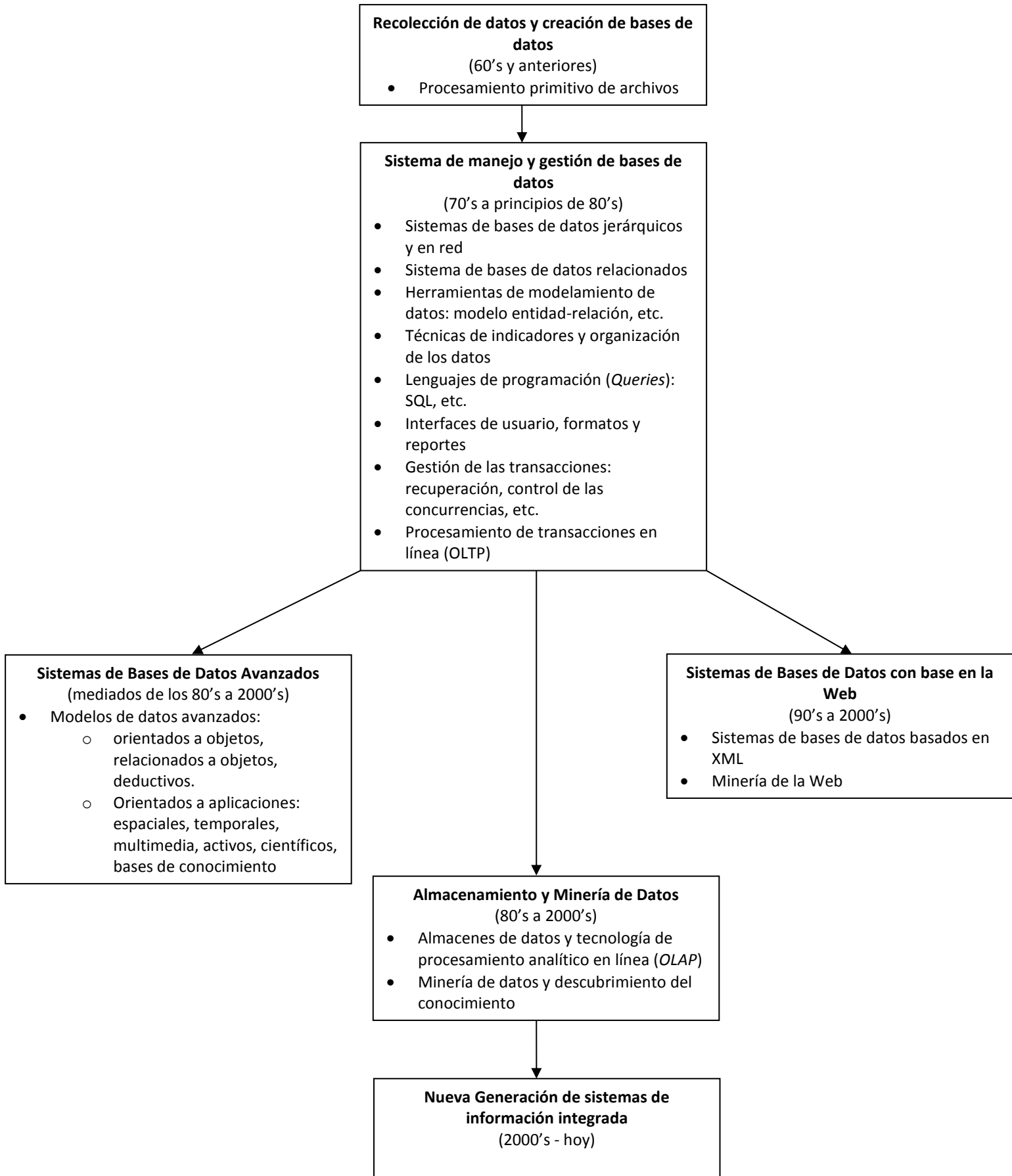


Fig. II.1 Evolución de la Tecnología

La abundancia de datos, sumada a la necesidad de herramientas de análisis de datos poderosas fue descrita como una situación de “rico en datos pero pobre en información”. El rápido crecimiento en la cantidad de la información recopilada y almacenada en grandes y numerosas bases de datos, excedió las habilidades humanas para su comprensión sin utilizar herramientas poderosas. Como resultado, los datos recopilados en grandes bases de datos se convirtieron en “tumbas de datos” y como consecuencia, las decisiones importantes frecuentemente eran tomadas basándose en la intuición de las personas encargadas de la toma de decisiones y no en los datos ricos en información almacenados en las bases de datos, simplemente porque no se contaba con las herramientas para extraer el conocimiento valioso que se encontraba depositado en las grandes cantidades de datos.

Hoy en día aunado a los problemas en la toma de decisiones que han aumentado en términos de número de clientes, variedad de productos, penetración de los canales de mercado, velocidad en el comercio, deserción, fraude, etc., se suma la reducción del tiempo disponible para la deliberación. Las técnicas de minería de datos pueden ser utilizadas para aprender acerca de los factores que soportan una decisión y construyen una aplicación que utiliza dichos factores para ayudar a la empresa a tomar esas decisiones de una manera objetiva y consistente.

Todo negocio que se jacte de maduro padece algunos problemas que parecen sin solución: juicios difíciles acerca de los clientes, asignación de los recursos, estrategia de negocios u organización. Estos constituyen el cinco por ciento de los problemas que son los principales objetivos de la minería de datos.

A medida que los directivos y gerentes crecen en experiencia, desarrollan y refinan los métodos para la toma de decisiones que mejor se adaptan a su empresa. Estos métodos formalizan conocimiento importante acerca de cómo opera su empresa. Las personas sobresalientes en la toma de decisiones son exitosas, frecuentemente, debido al conocimiento que poseen.

A través de la aplicación de la minería de datos y otras técnicas inteligentes de software, se pueden utilizar las computadoras para descubrir, almacenar y aplicar el conocimiento. Dado que el conocimiento es el corazón de la toma de decisiones efectiva, las técnicas de minería de datos para descubrir y explotar el conocimiento pueden ayudar a las personas en muchos aspectos del manejo de la empresa.

El conocimiento de la empresa se puede dividir en tres categorías:

- Conocimiento sobre el pasado, que es estable, voluminoso y relativamente preciso.
- Conocimiento sobre el presente, que es inestable, compacto y relativamente impreciso.
- Conocimiento sobre el futuro (predicciones o estimaciones), que es hipotético.

La minería de datos también provee las herramientas y técnicas esenciales para la optimización de las relaciones con los clientes y permite a las organizaciones usar sus datos activamente para lograr impactos potentes y no sólo el almacenamiento pasivo de los mismos.

Hoy en día los analistas requieren asistencia automática para enfocar su atención en los aspectos de producción de información de su fuente de datos. La necesitan para ayudarse a descubrir los puntos clave y las interrelaciones que serán valiosas en investigaciones futuras. El común denominador para todos los usuarios, es la necesidad de maximizar la cantidad de información útil que es extraída mientras que minimizan la pérdida de productividad que resulta del proceso de recuperación de la información y que es mermado por la prohibición en los accesos, las dificultades de comunicación y la recuperación de datos sin sentido, incorrectos o corruptos. Para satisfacer las necesidades de los analistas, los sistemas de minería de datos pueden ser diseñados para representar asociaciones en grandes conjuntos de datos, extraer y transmitir el material a través de una red de usuarios distribuidos en diversas locaciones, y analizar, interpretar y desplegar la información en una gran variedad de maneras significativas.

La minería de datos provee métodos automáticos para descubrir patrones en los datos. Las aplicaciones de minería de datos absorben y correlacionan los datos de una manera comprensible, contrario a lo que hacen los analistas humanos quienes son guiados por sus instintos en una manera secuencial. Claro está que, para alcanzar la totalidad de su potencial, la minería de datos debe ser un esfuerzo de colaboración entre la máquina y el analista.

Las organizaciones que más comúnmente se benefician de la minería de datos son las que:

- Cuentan con un gran volumen de datos;
- Tienen comunidades de colaboradores que necesitan entender los datos, pero no están preparados en temas de estadística;
- Tienen datos organizacionales complejos en su naturaleza, es decir, detallados y multifacéticos, con complejas relaciones entre ellos; y
- Operan en mercados competitivos

Actualmente las herramientas de minería de datos proveen mayores capacidades de análisis y desempeño dentro de un marco de fácil uso y acceso. Los actuales ambientes de minería de datos hacen uso extensivo de gráficos tanto para el usuario como para la representación de los resultados.

El objetivo de las herramientas de minería de datos ya no está enfocado al relativamente pequeño mercado de especialistas empleados en operaciones e investigación de mercado. Se están volviendo progresivamente más fáciles para el uso de los “colaboradores del conocimiento” como analistas financieros y de mercado. El conocimiento adquirido a través de las actividades de minería de datos puede ser capturado por profesionales de la computación e implementado dentro de aplicaciones en los sistemas para servicio de un amplio espectro de usuarios de la aplicación, desde ejecutivos de un call center hasta Jefes de Tecnología de la información.

Por otro lado, las bases de datos facilitan el análisis manual y la elaboración de reportes, pero no hacen pronósticos. Las técnicas de minería de datos pueden ser utilizadas para construir modelos sobre la experiencia anterior del negocio que pueda ser aplicada a conseguir beneficios en el futuro.

La minería de datos es utilizada con mayor frecuencia para ayudar al descubrir relaciones, realizar elecciones, hacer predicciones y mejorar procesos.

- II.1.1 Descubrir relaciones:** Este tipo de uso para la minería de datos, se basa en la observación de eventos recurrentes y se conoce en mercadotecnia como *Market Basket Analysis* aunque su nombre técnico es **análisis de enlaces o conexiones**. Este tipo de análisis se aplica claramente en las ventas directas. Opciones de compra de un producto se ofrecen a los consumidores que adquieren otros productos debido a que están frecuentemente ligados a ellos. Históricamente se ha demostrado que existe una gran propensión a realizar estas compras extras. Este análisis se describirá más a fondo en el capítulo IV.4
- II.1.2 Realizar elecciones:** Una decisión es una elección entre varias alternativas disponibles en el momento. En el mundo de los negocios, los encargados de la toma de decisiones tienen que elegir entre materiales caros o baratos, transacciones válidas o fraudulentas y hasta entre buenos y malos clientes. En estos casos, la minería de datos puede ser utilizada para evaluar los datos disponibles para tomar la decisión adecuada, aplicar una clasificación técnica y sugerir o priorizar la mejor manera o maneras entre distintas alternativas. Este tipo de conocimiento puede facilitar las ventas proactivas y que los clientes las tengan en cuenta más como un servicio que como una molestia.
- II.1.3 Realizar predicciones:** Una predicción es una elección entre varias alternativas disponibles en el futuro. Las predicciones normalmente anticipan comportamientos futuros de los clientes de manera individual, por ejemplo si un cliente pagará una deuda o no. Las predicciones también pueden anticipar el comportamiento a un nivel de mercado. Los modelos predictivos pueden ser utilizados para extrapolar la historia pasada en combinación con las condiciones actuales para predecir condiciones futuras.
- II.1.4 Mejorar procesos:** Los negocios exitosos han logrado optimizar los procesos de interacción entre las partes de un complejo sistema en el que interviene la relación con los proveedores, la infraestructura en las comunicaciones y hasta las presiones en cuanto a la regulación y el mercado. La efectividad en sus procesos puede llevar a una empresa al éxito o al fracaso. La minería de datos puede ser utilizada para revelar aspectos en el proceso de negocios que no son óptimos y estimar los efectos de llevar a cabo modificaciones que sean propuestas a estos procesos.

II.2 TIPOS DE DATOS NECESARIOS PARA LA APLICACIÓN DE LA MINERÍA DE DATOS

En principio, la minería de datos debería ser aplicable en cualquier repositorio de información, lo cual incluye bases de datos relacionales, almacenes de datos como se vio en 1.3, bases de datos transaccionales, sistemas de bases de datos avanzados como son las bases de datos orientadas a objetos, bases de datos de relación de objetos y bases de datos orientadas a aplicaciones (bases de datos espaciales, bases de datos de series de tiempo, bases de datos de texto y bases de datos multimedia), así como archivos planos y la *World Wide Web*. Las técnicas de minería de los datos difieren de acuerdo a cada tipo de repositorios. A continuación se explican brevemente algunos de estos tipos de repositorios de información:

- II.2.1 Bases de datos relacionales:** Un sistema de bases de datos, también conocido como sistema de manejo de bases de datos (*DBMS* por sus siglas en inglés), consiste en un conjunto de datos interrelacionados conocido como base de datos y un conjunto de programas de software para manejar y acceder a dichos datos. Una base de datos relacional es un conjunto de tablas, cada una de las cuales tiene un nombre único asignado. Cada tabla consiste de una serie de atributos (columnas o campos) y usualmente almacenan una gran cantidad de listas ordenadas de elementos o tuplas (registros, filas o coordenadas). Cada una de estas listas ordenadas en una tabla relacionada, representa un producto que se identifica con una única clave y que se describe por una serie de atributos. Se puede acceder a los datos relacionados a través de consultas (*queries*) desarrolladas en un lenguaje de programación relacionado como puede ser SQL. Cuando se aplica la minería de datos en este tipo de bases de datos, es posible buscar tendencias o patrones en los datos.
- II.2.2 Bases de datos transaccionales:** Consisten en un archivo donde cada registro representa una transacción. Una transacción típicamente incluye un único número identificador de la transacción (*trans_ID*) y un listado de los objetos que construyeron dicha transacción (como los objetos comprados en una tienda). La base de datos transaccional puede tener tablas adicionales asociadas a ella con información con respecto a la venta como la fecha de la transacción, el número identificador del cliente, el número identificador del vendedor y de la sucursal donde se originó la venta, etc.
- II.2.3 Sistemas de bases de datos avanzados:** Con la aparición de nuevas aplicaciones en lo que a bases de datos se refiere, se ha hecho necesario el desarrollo de nuevos y más avanzados sistemas de manejo de las bases de datos. Las nuevas aplicaciones en cuanto a bases de datos incluyen el manejo de datos espaciales (como los mapas), datos de ingeniería del diseño (como el diseño de edificios, componentes de los sistemas o circuitos integrados), hipertextos y datos multimedia (textos, imágenes, video y audio), datos relacionados con el tiempo (registros históricos o datos de intercambio de materiales) y la *World Wide Web* (un inmenso repositorio de información disponible en Internet). Estas aplicaciones requieren estructuras de datos eficientes y métodos a escala para el manejo de estructuras de objetos

complejas, registros de longitud variable, datos semi estructurados o no estructurados, texto y datos multimedia, y esquemas de bases de datos con estructuras complejas y cambios dinámicos.

II.2.3.1 Bases de datos orientadas a objetos: Basadas en la programación orientada a objetos, donde en términos generales, cada entidad es considerada como un objeto. Tales objetos pueden ser empleados, clientes o productos. Los datos y el código relacionado a un objeto están *encapsulados* en una sola unidad. Cada objeto se asocia con lo siguiente:

- Un conjunto de variables que describen los objetos (atributos).
- Un conjunto de mensajes que el objeto utiliza para comunicarse con otros objetos o con el resto del sistema.
- Un conjunto de métodos, donde cada método contiene el código para implementar un mensaje.

Los objetos que comparten un conjunto de propiedades similares pueden agruparse en una **clase**. Cada objeto es una **instancia** de dicha clase. Las clases pueden organizarse en **jerarquías** de clases o subclases.

II.2.3.2 Bases de datos relacional - objeto: Construidas basándose en el modelo de datos de relación de objetos. Este modelo extiende el modelo de relación de datos proporcionando un tipo de datos valioso para manejar objetos complejos; la orientación a objetos se realiza añadiendo la posibilidad de manejar tipos de datos complejos, jerarquías y herencia en los objetos.

II.2.4 Bases de datos espaciales: Contienen información relacionada con el espacio. Estas bases de datos incluyen bases de datos geográficas (mapas) y bases de datos médicas y de imágenes satelitales. Las bases de datos geográficas tienen un gran número de aplicaciones que van desde la planeación forestal y ecológica hasta proveer información de los servicios públicos como la localización de los cables eléctricos y de teléfono. También son utilizadas en los sistemas de navegación de los vehículos. La minería de datos aplicada en este tipo de bases de datos puede ayudar a descubrir patrones que describan las características de las casas localizadas cerca de un tipo de locación específico, como un parque. Otros patrones pueden describir el clima en áreas montañosas localizadas en varias altitudes, o describir el cambio en cuanto a la tendencia de las tasas de pobreza metropolitana basándose en la distancia a la que se encuentran las ciudades de las carreteras más importantes.

II.2.5 Bases de datos temporales y de series de tiempo: Almacenan datos relacionados con el tiempo. Una base de datos temporal usualmente almacena datos relacionados que incluyen atributos relacionados con el tiempo. Una base de datos de series de tiempo almacena secuencias de valores que cambian con el tiempo, como los datos recopilados al analizar el intercambio de materiales. La minería de datos, en estos casos, se puede

utilizar para encontrar las características en la evolución de un objeto o la tendencia de cambios en los objetos de la base de datos.

II.2.6 Bases de datos de texto y multimedia: Las bases de datos de texto contienen descripciones escritas para los objetos. Este tipo de descripciones normalmente no son simples palabras claves sino largas oraciones o párrafos, como son las especificaciones de un producto, reportes de errores, mensajes de alerta, reportes resumen, notas, u otros documentos. Estas bases de datos pueden no estar estructuradas o ser semi estructuradas. La minería de datos ayuda a descubrir descripciones generales de clases de objetos, así como asociaciones por palabra clave o contenido, y el comportamiento grupal de objetos de texto. Para realizar esto, métodos de minería de datos estándar necesitan integrarse con técnicas de retorno de información y con la construcción o uso de jerarquías específicas para datos de texto (como diccionarios), así como sistemas de clasificación de términos orientados a disciplinas (como medicina, química o economía).

Las bases de datos multimedia almacenan datos de imagen, video y audio. Se utilizan en aplicaciones como sistemas de correo de voz, de video *on-demand*, la *World Wide Web* e interfaces de usuario basadas en comandos por voz que reconocen instrucciones habladas. Para la minería de datos en este tipo de bases de datos, se necesitan integrar técnicas de almacenamiento y búsqueda con los métodos de minería de datos estándares.

II.2.7 La World Wide Web: La red o WWW por sus siglas en inglés y los sistemas de distribución de información como Yahoo! o Google, proveen valiosos servicios de información en línea a lo largo del mundo, donde los objetos de datos están ligados para facilitar el acceso interactivo. Este tipo de sistemas siempre han proporcionado amplias oportunidades para la minería de datos y es por eso que hoy en día son tan populares. Como ejemplo tenemos los anuncios en las páginas de Internet que se basan en los patrones de visita de los usuarios.

II.3 FUNCIONALIDADES DE LA MINERÍA DE DATOS

Las funcionalidades de la minería de datos ayudan a especificar el tipo de patrones a ser encontrados en las tareas efectuadas, que bien pueden clasificarse en dos categorías: descriptivas y predictivas. Las tareas descriptivas caracterizan las propiedades generales de los datos en la base de datos. Las tareas predictivas realizan inferencias en los datos actuales con el fin de hacer predicciones.

Los sistemas de minería de datos deben permitir su utilización en tipos de patrones múltiples que se adecúen a diferentes aplicaciones o expectativas de los usuarios. Más aún, los sistemas de minería de datos deben ser capaces de descubrir patrones a distintos niveles de abstracción o vistas y deben permitir a los usuarios especificar *hints* para guiar o enfocar la búsqueda de patrones interesantes.

A continuación se describen las funcionalidades de la minería de datos y el tipo de patrones que pueden ayudar a descubrir.

II.3.1 Descripción de Concepto/Clase: Caracterización y Discriminación: Los datos pueden ser asociados con clases o conceptos, por ejemplo, en una tienda de aparatos electrónicos, las clases de productos pueden incluir *computadoras* e *impresoras* y los conceptos de clientes pueden incluir *grandes compradores* y *compradores promedio*. Es útil pues, describir las clases y conceptos individuales con términos resumidos, concisos y precisos. Estas descripciones de clases y conceptos se llaman descripción de concepto/clase y pueden obtenerse a través de (1) *caracterización de los datos*, mediante el resumen de los datos de la clase a ser estudiada (clase objetivo), (2) *discriminación de los datos*, mediante la comparación de la clase objetivo con una o un conjunto de clases comparativas (clases contraste), o (3) ambas, caracterización y discriminación de los datos.

La caracterización de los datos es un resumen de las características generales o características específicas de una clase objetivo de datos. Los datos correspondientes a la clase especificada por el usuario son típicamente recabados por una consulta en la base de datos. El resultado de la caracterización de los datos puede ser presentado de varias formas como son los gráficos de *pie*, gráficos de barras, curvas, cubos de datos multidimensionales y tablas multidimensionales, incluyendo tablas cruzadas.

La discriminación de los datos es una comparación de las características generales de los objetos de datos de una clase objetivo con las características generales de los objetos de una o un conjunto de clases contrastes. Las clases objetivo y contraste pueden ser especificadas por el usuario y los datos obtenidos a través de *queries* en las bases de datos. La presentación de los resultados de la discriminación de datos son similares a los de la

caracterización sólo que se deben incluir medidas comparativas que ayuden a distinguir entre clases objetivos y contrastes.

- II.3.2 Análisis de asociación o asociativo:** Este tipo de análisis se basa en el descubrimiento de reglas de asociación mostrando condiciones de atributo-valor que ocurren frecuentemente juntos en un determinado conjunto de datos. Se utiliza ampliamente para los análisis de *Market Basket* o de enlaces o conexiones definidos en el punto II.1.1 y de transacciones.

Las reglas de asociación son de la forma $X \Rightarrow Y$, esto es:

$$A_1 \dots A_m \longrightarrow B_1 \dots B_n, \text{ donde}$$

A_i para $i \in \{1, \dots, m\}$ y B_j para $j \in \{1, \dots, n\}$ son valores de atributos pares.

La regla de asociación $X \Rightarrow Y$ se interpreta como “las listas ordenadas de objetos o tuplas de la base de datos, que satisfacen las condiciones en X y que son también propensas a satisfacer las condiciones en Y ”

- II.3.3 Clasificación y Predicción:** Clasificar es el proceso de encontrar un conjunto de modelos (o funciones) que describan y distingan clases de datos o conceptos, con el fin de ser capaz de utilizar dicho modelo para predecir la clase de objetos cuya etiqueta de clase es desconocida. El modelo está basado en el análisis de un conjunto de objetos de datos cuya etiqueta de clase es conocida.

Los modelos de clasificación pueden ser representados en varias formas como son las reglas de clasificación SI-ENTONCES, árboles de decisión o fórmulas matemáticas.

La clasificación puede ser utilizada para predecir la etiqueta de clase de los objetos de datos. Sin embargo, en muchas aplicaciones, los usuarios desearían predecir valores de los datos faltantes o no disponibles más que las etiquetas de las clases. Este es usualmente el caso en el que los valores predichos son datos numéricos conocidos como predicciones. La predicción también ayuda a la identificación de tendencias de distribución basadas en datos actualmente disponibles.

- II.3.4 Análisis Cluster o De Grupos (definición):** El análisis *cluster*, a diferencia de la clasificación y la predicción, analiza objetos de datos sin consultar una etiqueta de clase conocida. En general, las etiquetas de clase no están presentes en los datos preparatorios simplemente porque no son conocidas al comienzo. La agrupación puede ser utilizada para generar dichas etiquetas. Los objetos son agrupados mediante el principio de maximizar la similitud dentro de una clase y minimizar la similitud entre clases. Cada grupo puede ser visto como una clase de objetos y de ahí derivar las reglas.

- II.3.5 Análisis Outlier o de Partes Aisladas:** Una base de datos puede contener objetos de datos que no cumplen con el comportamiento general o el modelo

de los datos. Estos datos son conocidos como partes aisladas. La mayoría de los métodos de minería de datos descartan este tipo de partes aisladas como excepciones. Sin embargo, en algunas aplicaciones como la detección de fraudes, estos eventos extraños pueden ser aún más interesantes que los eventos que ocurren con regularidad.

Estas partes aisladas pueden ser detectados utilizando pruebas estadísticas que asuman un modelo de distribución o probabilidad para los datos, o usando medidas de distancia donde los objetos que tienen una distancia considerable de un grupo se consideran partes aisladas; mejor aún, los métodos basados en la desviación de datos identifican las partes aisladas examinando las diferencias entre las principales características de los objetos en un grupo.

II.3.6 Análisis de Evolución: Describe y modela regularidades y tendencias para objetos cuyo comportamiento cambia a lo largo del tiempo. En este tipo de análisis cabe el análisis de series de tiempo, la igualación de patrones de secuencia o periodicidad y análisis de datos basado en similitudes.

La minería de datos puede arrojar miles o hasta millones de patrones diferentes con respecto a una base de datos. Los patrones que deben considerarse como objetos de estudio o valiosos son los patrones interesantes, es decir, fácilmente comprensibles, válidos en datos nuevos o de prueba con cierto grado de precisión, que sean potencialmente útiles y originales, así como que validen la hipótesis que el usuario buscaba confirmar.

Por otro lado un algoritmo o sistema de minería de datos rara vez genera todos los patrones interesantes que se buscan por lo que se deben poner algunas restricciones y medidas en cuanto a interés se refiere para enfocar la búsqueda.

Finalmente, los algoritmos o sistemas de minería de datos no sólo generan patrones interesantes por lo que se debe trabajar constantemente en la optimización de dichos sistemas.

II.4 EL PROCESO DE LA MINERÍA DE DATOS

La minería de datos es un proceso que, en términos generales, comienza con una serie de *datos* y termina con un cúmulo de información interesante llamada *conocimiento*. Para entender mejor este proceso, vale la pena definir algunos conceptos previos como son el descubrimiento y explotación de la información y los llamados modelos.

II.4.1 EL DESCUBRIMIENTO Y LA EXPLOTACIÓN

Durante la etapa del descubrimiento se detectan y caracterizan formalmente patrones significativos en los datos (modelos descriptivos); durante la etapa de explotación se utilizan tales patrones significativos para crear aplicaciones útiles en la toma de decisiones de cada negocio (modelos predictivos).

Varias de las herramientas y técnicas que se utilizan dentro del proceso de minería de datos y que son promovidas por el mercado de ventas de las aplicaciones de Inteligencia de Negocios (*BI*), pueden ser utilizadas tanto para la etapa de descubrimiento como para la de explotación. Cada vez con más frecuencia, los analistas quieren engancharse en ambas etapas utilizando herramientas y métodos de minería de datos y pasando de una a otra durante todo el proyecto. Lo que siempre se debe tener claro, es la diferencia entre una y otra etapa para ser capaces de discernir entre ambas y llevar a cabo la tarea que sea necesaria para cada proyecto.

La meta y los objetivos de cada empresa es lo que debe determinar si tanto el descubrimiento como la explotación son los objetivos para cada proyecto de minería de datos en particular. Típicamente, los analistas e investigadores quieren descubrir, mientras que los gerentes desean explotar. El descubrimiento es un prerrequisito para la explotación, sin embargo, en algunas ocasiones, se cuenta con el conocimiento suficiente de los datos que permite la explotación (actividades de modelamiento) sin necesidad de un profundo esfuerzo en el descubrimiento.

II.4.2 MODELOS

Para entender mejor el proceso en general, es conveniente apoyarse en el significado de *ontología = teoría del ser*. Para este caso en particular, la ontología de un conjunto de datos es un esquema que provee una presentación consistente, coherente y unificada de los datos en contexto, es decir, una interpretación de los datos que revela su significado explicando sus características. Estas representaciones son los llamados modelos. Los modelos representan ontologías de varias maneras: como conjuntos consistentes de ecuaciones que describen la relación entre los datos y patrones (modelos matemáticos), como una recopilación de leyes empíricas o principios (modelos científicos), etc.

Si los modelos describen patrones en los datos, son llamados modelos descriptivos. Si los modelos procesan patrones en los datos con el propósito de predecir resultados, son llamados modelos predictivos.

A continuación se describen los pasos y actividades que conforman el proceso de la minería de datos y entre paréntesis, como porcentaje, el tiempo que se debe invertir en cada uno:

- **Paso 1 - Definición del problema (5%):** Tiene como propósito la creación de una lista de análisis a desarrollar, preguntas que investigar e hipótesis que probar, es decir, formular un objetivo claro que sea cumplido. En un esfuerzo de explotación, el éxito usualmente significa el desarrollo de un modelo predictivo efectivo que nos ayude a prevenir situaciones de riesgo y mejorar las tareas para obtener mejores resultados.
- **Paso 2 - Evaluación de los datos (20%):** Su propósito es obtener una representación efectiva del dominio (ingesta y formato de datos). Este paso incluye cuatro actividades: aleatoriedad, muestreo, partición y *hosteo*.
La **aleatoriedad** implica establecer un esquema que asegure que la selección de los conjuntos de datos a ser utilizados para el análisis serán muestras aleatorias.
El **muestreo** implica la selección de un subconjunto de datos (una muestra) para ser utilizada en el análisis. Debe ser una muestra representativa.
La **partición** implica la selección de un número apropiado de muestras para aspectos del análisis en particular.
Finalmente, el **hosteo** se refiere a cómo los datos serán almacenados físicamente y cómo se accederá a ellos. Esto se puede resolver construyendo bases o almacenes de datos.
- **Paso 3 – Extracción y mejora de los datos (35%):** Su propósito es producir un conjunto de datos representativo, reproducible y confiable (preciso y completo) listo para la minería de datos.
- **Paso 4a - Elaboración de un prototipo del plan (20%):** Este paso tiene como propósito el desarrollo de hipótesis de trabajo para el análisis basadas en las metas de la empresa para posteriormente construir un plan para probarlas.
- **Paso 4b - Elaboración de un prototipo de desarrollo de modelos (20%):** Su propósito, como su nombre lo indica, es el de desarrollar modelos: descriptivos, predictivos o ambos. Este paso constituye lo que la mayoría de la gente llama “minería de datos” y es en este paso en el que el analista trabaja con los datos.
- **Paso 5 - Evaluación del modelo (3%):** Tiene como propósito evaluar los resultados del último ciclo de elaboración de prototipos. Durante este paso, el analista y gerente de programación toman material sobre el progreso en la etapa de descubrimiento y explotación y documentan los éxitos y fracasos.

- **Paso 6 – Implementación (15%):** En este paso se incluye la documentación breve y por escrito del esfuerzo para la presentación del conocimiento encontrado así como la terminación de las interfaces de usuario, la documentación técnica y de usuario por escrito, la población de las estructuras de datos operacionales y la creación e instalación del producto final.
- **Paso 7 - Evaluación del retorno de inversión (ROI) (post proyecto):** El impacto financiero del proceso de minería de datos se calcula y analiza. Se debe asegurar que para estos cálculos, se incluya los costos de cualquier error que tenga el modelo.

Los pasos anteriores son aplicados de acuerdo a la *metodología de prototipos rápidos* y son lo suficientemente flexibles tanto en orden como en contenido, para que puedan ser reorganizados, de acuerdo a objetivos y repetidos si fuera necesario, con excepción de los pasos 6 y 7 que deben llevarse a cabo secuencialmente y siempre como los últimos pasos del proceso.

II.5 LAS MATEMÁTICAS EN LA MINERÍA DE DATOS

Las bases de los métodos de minería de datos en lo que se refiere a descubrimiento y explotación, se encuentran en el marco de las matemáticas, que consiste en técnicas del álgebra lineal, probabilidad y estadística en su mayoría.

La minería de datos avanzada requiere que los datos se expresen en una forma que facilite la representación y el análisis. Como las aplicaciones de minería de datos están íntimamente ligadas con las matemáticas, muchas técnicas de minería de datos funcionan exclusivamente de forma numérica. Las matemáticas proveen el marco ideal en el que la minería de datos se debe conducir y éste es el “espacio Euclidiano”: $R^1, R^2, R^3, \dots, R^n$. *Cada problema de minería de datos que utilice datos numéricos puede ser formulado en su propio espacio Euclidiano.* El plasmar los valores de los datos en un espacio de dimensión N ayuda a representarlos como un conjunto de puntos en un espacio Euclidiano de dimensión N .

La estadística es vital para un proceso de minería de datos exitoso. Por un lado, la minería de datos busca patrones, por el otro, la estadística provee formalismos matemáticos para cuantificar la presencia y el significado de dichos patrones como lo puede hacer un simple histograma.

La estadística se utiliza para:

- Estimar la complejidad del problema de minería de datos;
- Sugerir qué tipo de técnicas de minería de datos son más probables para ser utilizadas;
- e
- Identificar los campos en los datos que contengan la mejor información

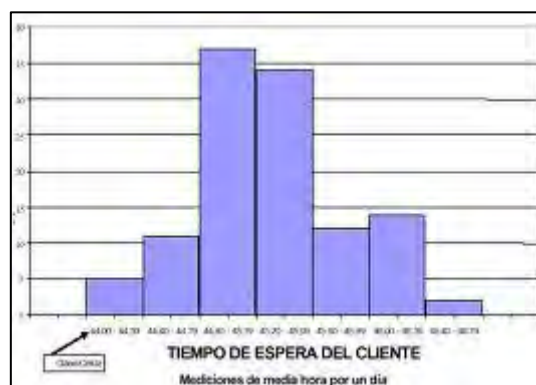


Fig. II.2 Histograma

También las herramientas de probabilidad son básicas para obtener mejores resultados en el proceso de la minería de datos. Las distribuciones de probabilidad más importantes para las aplicaciones de minería de datos son la distribución normal y la distribución uniforme.

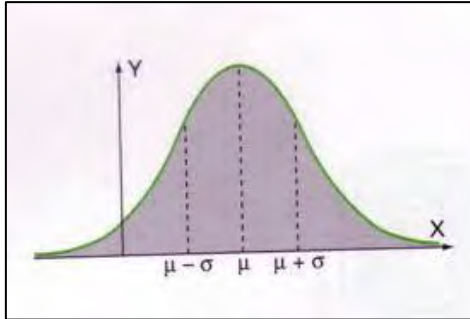


Fig. II.3 Gráfica de una distribución normal

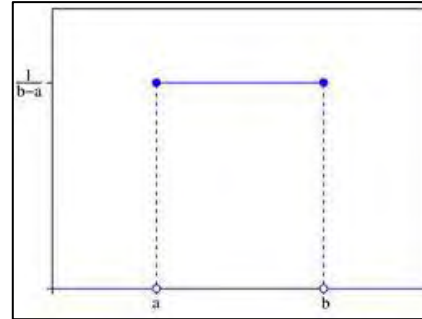


Fig. II.4 Gráfica de una distribución uniforme

Por otro lado, la desviación estándar σ es utilizada de tres maneras durante el proceso de minería de datos: para describir los datos, para verificar la limpieza de los datos y para normalizar los datos.

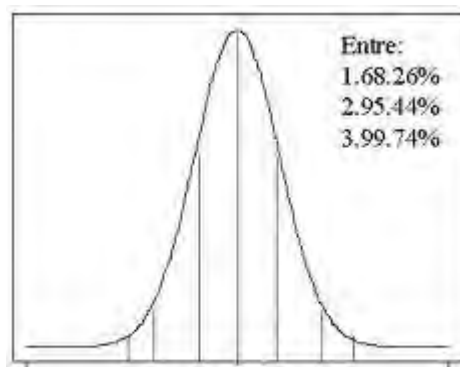


Fig. II.5 Gráfica de desviación estándar

Como ya también se explicó en el punto II.3.4, tenemos los *clusters* o grupos, que son también conceptos fundamentales dentro del espacio Euclidiano.

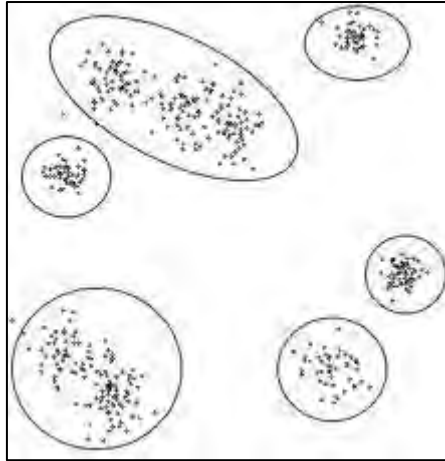


Fig. II.6 Clusters

Para tener éxito en la aplicación de los métodos matemáticos en la minería de datos, se pueden convertir las características nominativas de una población (número telefónico, código postal, género, mes de nacimiento, ciudad de origen, etc.), en características numéricas, a través de la codificación numérica, es decir, la asignación de un valor numérico significativo a un símbolo que no es numérico. Esto se puede hacer si el analista, por ejemplo, convierte en una medida numérica importante para el negocio, dichos atributos nominales. Por ejemplo, en una empresa de mensajería, a los habitantes de México D.F. se les asignará un número 1.00 puesto que el valor del envío es de 100 pesos mientras que a los habitantes de Chihuahua se les asignará el 2.40 puesto que el valor del envío es de 240 pesos.

De igual manera, se puede etiquetar un registro con una característica nominal a partir de una característica numérica determinada. Por ejemplo, los consumidores que gastan 1,000 pesos al mes se etiquetan en la categoría "A", los que gastan 1,000 pesos a la quincena en la "AA", los que gastan 1,000 pesos a la semana caen en la categoría "AAA", etc.

II.6 TÉCNICAS DE MINERÍA DE DATOS

A continuación se describirán algunas de las técnicas de minería de datos más representativas que ayudan y soportan la obtención del conocimiento: Análisis *cluster* o de grupos, análisis de relaciones y visualización de los datos así como los llamados Modelos predictivos.

II.6.1 Análisis Cluster (explicación): Este tipo de análisis, también llamado “de grupos” se basa en el proceso de agrupar un conjunto de objetos o datos similares en clases o grupos. Cada grupo o *cluster*, es una colección de objetos de datos que son similares a otros dentro del mismo grupo y distintos a los objetos en otros grupos. En los negocios, por ejemplo, este tipo de técnica puede ayudar a los distintos mercados a determinar diferentes grupos en sus bases de clientes y caracterizar dichos grupos de acuerdo a sus patrones de compra.

Como una rama de la estadística, el análisis *cluster* ha sido desarrollado por muchos años, enfocándose principalmente en el *análisis cluster basado en distancias*. Las herramientas de análisis *cluster* están basadas en métodos como el k-medias, o el *k-medoids* que han dado como resultado la creación de software para análisis estadístico como SPSS y SAS.

Para llevar a cabo este análisis para la minería de datos se deben cumplir algunos requerimientos:

- Escalable: Que funcione tanto para pequeñas muestras como para grandes conjuntos de objetos o datos.
- Habilidad para tratar con diferentes tipos de atributos: datos numéricos, datos binarios, nominativos, ordinales o mixtos.
- Que ayuden a descubrir *clusters* con forma arbitraria: Desde esféricos hasta con formas irregulares.
- Que contengan el mínimo dominio del conocimiento para determinar parámetros de alimentación de información.
- Con la habilidad para tratar con datos “ruidosos”: Datos aislados, faltantes, desconocidos o erróneos.
- No sensibles al orden de alimentación de los datos.
- Manejo de N dimensiones: Cuando N es mayor a 3.
- Manejo de restricciones dadas.
- Resultados interpretables y útiles

Supongamos que el conjunto de datos al que será aplicado el análisis contiene n objetos, que pueden representar personas, documentos, clientes, países, etc., entonces, se puede aplicar el análisis *cluster* sobre dos tipos de datos:

II.6.1.1 Matriz de datos: Representa n objetos que pueden ser personas, con p variables (también llamadas medidas o atributos), como pueden ser la edad, altura, peso, género, etc. La estructura es de la siguiente forma:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

II.6.1.2 Matriz de disimilitud: Almacena una colección de proximidades disponibles para todos los pares de n objetos. Se representa de la siguiente forma:

Group	1	2	3	4	5	6	7	8
1	0							
2	4.15	0						
3	11.02	15.01	0					
4	7.16	3.03	18.02	0				
5	43.72	47.49	32.80	50.41	0			
6	54.37	58.23	43.36	61.19	11.12	0		
7	46.34	50.20	35.34	53.16	3.78	8.03	0	
8	55.42	59.27	44.42	62.23	12.05	1.12	9.08	0

Donde $d(i,j)$ es la diferencia o disimilitud medible entre los objetos i y j . En general, $d(i,j)$ es un número no negativo cercano a 0 donde los objetos i y j son altamente similares o “ceranos” el uno del otro, y se vuelve mayor en cuanto más difieren los objetos, además $d(i,j) = d(j,i)$ y $d(i,i) = 0$

Varios algoritmos del análisis *cluster* operan sobre la matriz de disimilitud por lo que se tienen algunas técnicas para evaluar dicha condición.

- Variables Escalares de Intervalo: Son medidas continuas aproximadas de una escala lineal como pueden ser el peso y la altura, las coordenadas de latitud y longitud, etc. Las unidades de medida utilizadas, frecuentemente afectan el análisis de *cluster*. Para evitar la dependencia en la elección de unidades de medida, los datos deben ser estandarizados con lo que se intenta dar a todas las variables un peso igual.

Para estandarizar dichas medidas, una opción es convertir las medidas originales en variables sin unidad. Dadas medidas para la variable f , esto puede realizarse de la siguiente manera:

a) Calcular la **desviación media** (media de las desviaciones absolutas) , s_f :

$$s_f = \frac{1}{n} (|x_{1f} - m_f| + |x_{2f} - m_f| + \dots + |x_{nf} - m_f|),$$

Donde x_{1f}, \dots, x_{nf} son n medidas de f, y m_f es la media de f, que es

$$m_f = \frac{1}{n} (x_{1f} + x_{2f} + \dots + x_{nf}).$$

b) Aplicar la **transformación z** o estandarización de los datos:

$$z_{if} = \frac{x_{if} - m_f}{s_f}$$

Una vez hecha la estandarización, la disimilitud (o similitud) entre objetos descritos por las variables escalares de intervalo se calcula de acuerdo a la distancia entre cada pareja de objetos. La medida de distancia más popular es la distancia Euclidiana:

$$d_{(i,j)} = \sqrt{|x_{i1} - x_{j1}|^2 + |x_{i2} - x_{j2}|^2 + \dots + |x_{ip} - x_{jp}|^2}$$

Donde $i = (x_{i1}, x_{i2}, \dots, x_{ip})$ y $j = (x_{j1}, x_{j2}, \dots, x_{jp})$ son dos objetos de datos de dimensión p .

Otra métrica es la distancia de Manhattan definida como:

$$d_{(i,j)} = |x_{i1} - x_{j1}| + |x_{i2} - x_{j2}| + \dots + |x_{ip} - x_{jp}|$$

Donde ambas cumplen con que:

1. $d_{(i,j)} \geq 0$: La distancia no es negativa
2. $d_{(i,i)} = 0$: La distancia de un objeto hasta sí mismo es 0
3. $d_{(i,j)} = d_{(j,i)}$: La distancia es una función simétrica
4. $d_{(i,j)} \leq d_{(i,h)} + d_{(h,j)}$: Ir directamente del punto objeto i al objeto j en el espacio no es mayor que pasar por otro objeto h (*Desigualdad del triángulo*).

II.6.2 Análisis de relaciones: Esta importante y poderosa técnica ayuda a descubrir información en conjuntos de datos grandes y complejos. Es una estrategia de la minería de datos para identificar eventos que ocurren juntos. Un evento en este sentido puede ser cualquier resultado que el analista quiera definir: el desgaste de los

clientes, comisión en los fraudes, ocurrencia de cierto tipo de venta o compra, etc. El objetivo de este análisis es encontrar indicadores comunes de un evento para que la oportunidad correspondiente pueda ser explotada. Algunos algoritmos para llevar a cabo el análisis de relaciones son:

- Asociaciones: Son grupos o “eventos” que regularmente ocurren juntos. Ejemplo: los compradores de un determinado producto son padres de familia.
- Patrones secuenciales: Son asociaciones que varían de maneras consistentes. El cambio puede ocurrir o no durante un periodo de tiempo. Ejemplo: los compradores de un determinado producto A y otro B, frecuentemente comprarán el producto C.
- Patrones de tiempo secuenciales: Son patrones secuenciales cuyos elementos ocurren en un marco de tiempo regular. Ejemplo: los compradores que asisten a las rebajas 2 veces en una temporada tienen el 80% de probabilidad de regresar una tercera vez.

La forma más simple de análisis de relaciones es la estratificación o la segmentación como se conoce en mercadotecnia. La estratificación divide la población en estratos para un propósito de análisis.

Una forma más sofisticada de este análisis puede ser aplicada cuando los datos son numéricos. El coeficiente de correlación de todos los posibles pares de datos puede ser calculado y colocado en una matriz de correlación. Esta técnica cuantifica la fuerza de la asociación y no requiere que el analista deduzca qué datos están siendo relacionados

II.6.3 Visualización de los datos: El objetivo de esta técnica es ayudar al analista a ganar intuición acerca de los datos que están siendo observados. En este tipo de técnicas se incluyen los planos, histogramas, árboles de relación, gráficos de barras y pie, tablas, etc., técnicas básicas de visualización que continúan siendo útiles. Por otro lado se tienen técnicas más actuales como son el *drill-down*, los hiper vínculos a fuentes de datos relacionadas, vistas múltiples y simultáneas de datos, zoom, las aplicaciones Java, íconos, etc. La visualización de datos ofrece varias ventajas como permitir a los analistas desarrollar su intuición sobre una población, ayuda a descubrir directamente reglas en un conjunto de datos determinado, detección de diferencias significativas entre clases, irregularidades en los datos, etc. Como limitaciones se encuentra que sólo se pueden visualizar los datos en pocas dimensiones cada vez y entre mayor es la cantidad de características a graficar, más difícil se vuelve la visualización de los resultados.

II.6.4 Modelos predictivos: Son los que utilizan los métodos de minería de datos para realizar una clasificación de patrones. Estos modelos pueden ser vistos a partir de un “problema de regresión”. En regresión, se construye un modelo que acepta datos conocidos específicos y a través de un cierto tipo de transformación matemática, produce el resultado deseado. Los datos conocidos que alimentan el modelo son características, y el resultado es la respuesta buscada: puntuación crediticia, decisión en un fraude, una predicción en el mercado. En estos casos, la minería de datos

considera el revisar la información histórica de una empresa o negocio que sea relevante para el futuro de la misma.

Uno de los modelos predictivos más simple es la **regresión polinomial** y se refiere al uso de un polinomio como modelo para un determinado conjunto de datos. Dado un grupo de parejas (x, y) en el plano, la regresión polinomial puede ser utilizada para encontrar una expresión que relacione un valor x dado con su correspondiente valor y . Siempre y cuando el problema sea sencillo y manejable, es decir, que los datos estén relacionados por una ecuación lineal o se ajusten adecuadamente a un modelo de probabilidad conocido. Normalmente en este caso se utiliza el estimador de *mínimos cuadrados* y entre mayor sea la complejidad del polinomio utilizado, es decir, en cuanto mayor sea el grado del polinomio utilizado, el modelo es más poderoso. Los modelos poderosos pueden capturar detalles finos de un conjunto de datos.

Otro modelo predictivo son los **Sistemas Expertos Basados en Conocimiento (KBES** por sus siglas en inglés), los cuales capturan conocimiento humano de nivel experto, en reglas que sean *ejecutables*. La unidad fundamental de conocimiento en los KBES es la regla, que usualmente tiene la forma de una construcción estructurada como lo es el argumento SI-ENTONCES-SI NO (IF-THEN-ELSE)

Finalmente se tienen las **Superficies de Decisión**, basadas en los valores de decisión que son aquellas características que dividen a los datos en subconjunto por clases. Las formas que constituyen los límites entre las clases son llamadas superficies de decisión. Todos los clasificadores operan creando, directa o indirectamente, superficies de decisión entre las clases. Normalmente en estos casos se aplica la regla de mínimos cuadrados para mover la superficie de decisión de tal manera que mejore su habilidad para discriminar entre clases. Este proceso se repite hasta que no haya ningún cambio en la posición de la línea que separa las clases.

II.7 GESTIÓN Y MANEJO DE LA MINERÍA DE DATOS

Para la correcta gestión de la minería de datos, es importante desarrollar todo un proyecto dentro de la empresa, que involucre a varios integrantes de la misma y que esté bien planeado desde un inicio, con una meta bien definida y siempre con un objetivo de mayor ganancia con menor inversión de recursos.

Un proyecto de minería de datos considera las siguientes fases para alcanzar el objetivo deseado:

II.7.1 Concepto de entrega del proyecto: Es la clave de éxito para iniciar un proyecto y se refiere al momento en que se establece la meta final o visión que se tiene del mismo. Se deben incluir resultados tangibles y medibles y determinar desde un principio con los clientes, cuáles son los productos que se entregarán como resultado de cada proyecto de minería de datos.

II.7.2 Análisis del proyecto: Es en esta fase en la que el líder de proyecto descubrirá las respuestas a contestar: *qué* se quiere encontrar, *quiénes* serán los encargados de llevar a cabo el trabajo, *dónde* se encuentran los datos a estudiar, *cuándo* se llevarán a cabo las tareas que conforman el proyecto, *por qué* se lleva a cabo dicho proyecto, *cómo* se llevará a cabo y, para los niveles directivos, *cuánto* costará el llevarlo a cabo.

II.7.3 Reunión del equipo para el proyecto: La cantidad de personas que conformarán el equipo que llevará a cabo el proyecto dependerá directamente del tamaño del mismo, pero siempre se identifican cuatro roles básicos que participan:

- Líder científico, que aporta conocimientos avanzados y conceptos científicos a la solución de un problema, por ejemplo: un matemático, actuariólogo, etc.
- Líder técnico, que sirve como foco técnico del proyecto y hará énfasis en la metodología y los resultados, por ejemplo: un ingeniero.
- Líder de proyecto, maneja y conduce el proyecto de tal manera que se alcancen los objetivos técnicos y de negocio. Igualmente es el encargado de desarrollar el plan del proyecto y se asegura de la comunicación entre todas las partes que conforman el mismo.
- Experto en la materia, es el encargado de llevar por buen camino, las tareas e investigaciones de los demás colaboradores así como de señalar las debilidades y fortalezas de los resultados encontrados en un determinado proyecto.

De igual manera se debe tener en cuenta el lugar en el que se lleve a cabo el proyecto, asegurándose de tener acceso a las herramientas y datos necesarios para el logro de los objetivos. La participación de usuarios claves que conozcan la información y los datos, que de igual modo participan directamente a las actividades del día a día en una empresa, hacen que muchas veces no se pueda dedicar el cien por ciento de la atención a los proyectos por lo que conviene considerar que estos usuarios clave disminuyan su participación en la operación diaria de la empresa. Una buena manera

de logra esto, es tener un lugar especial donde se desarrolle el proyecto para evitar distracciones y enfocar el cien por ciento de los esfuerzos en la realización del mismo.

II.7.4 Metodología: Con el fin de evitar malos entendidos entre los integrantes de un proyecto, vale la pena establecer desde un inicio, la metodología de trabajo que se seguirá a lo largo del mismo, estableciendo las “reglas del juego” bajo las cuales se realizará el trabajo. La metodología puede incluir los tiempos, la determinación de reuniones de trabajo o resultados, los documentos entregables, firma de responsabilidades, etc.

II.7.5 Calendario del proyecto: Antes de arrancar el proyecto, se deben calendarizar las actividades que se llevarán a cabo, con el fin de trabajar con miras a cumplir los tiempos y no aumentar los costos resultantes de posibles retrasos. Los pasos del Proceso de Minería de Datos del punto II.4 se deben acotar en marcos de tiempo determinados.

II.7.6 Costo del proyecto: Se deben estimar siempre, los costos derivados de un proyecto de minería de datos, considerando la inversión previa en cuanto a software, equipo, aplicaciones, etc. así como estimaciones en gastos de viaje o gastos extraordinarios que pueda conllevar el proyecto. Son útiles las experiencias en cuanto a proyectos previos para evaluar los nuevos costos.

II.7.7 Entrega del proyecto: Es en este punto cuando fallan muchos proyectos que comenzaron y se desarrollaron de manera adecuada. Cuando el equipo ha pasado la parte compleja de búsqueda y desarrollo, tiene a relajar esfuerzos lo que hace que el proyecto no se concluya de manera óptima. Es aquí donde el líder de proyecto debe empujar al equipo entero a que redoble esfuerzos con el fin de entregar el trabajo terminado.

II.8 LA MINERÍA DE DATOS EN LA INDUSTRIA - APLICACIONES

En los mercados competitivos, existen numerosas áreas que pueden beneficiarse de utilizar técnicas de la minería de datos como son los modelos predictivos, por ejemplo selección de nichos, tiempo para el lanzamiento de un producto, objetivos de marketing.

La madurez generalizada en el desarrollo de la minería de datos en la industria está en función de la magnitud y el tipo de información disponible. Por ejemplo, la industria de servicios financieros (bancos, financieras, inversionistas, etc.) ha sido una de las pioneras en la utilización de técnicas de minería de datos. Por otro lado, el marketing directo es una de las aplicaciones más populares de la minería de datos dado que los resultados se encuentran listos para obtener una mejor respuesta de campaña y un mejor retorno de inversión. Las campañas de marketing directo van más allá de la industria de las ventas menores de mercancías, alcanzando los servicios al cliente. La deserción de los clientes es uno de los mayores problemas en muchas industrias. Generalmente se considera que es mejor, en términos de costos, el retener a los clientes existentes que el asegurar a los nuevos. Esto implica que es deseable identificar a los clientes con riesgo a cambiar su servicio a otro proveedor para ofrecerles incentivos que los hagan querer retener el servicio con el que ya cuenta. Los datos históricos de los clientes, frecuentemente contienen información que puede ser altamente valiosa para los especialistas en retención. Los datos históricos sostienen patrones de uso y otras características importantes de los clientes que, una vez descubiertas, pueden ser utilizadas para identificar a clientes satisfechos e insatisfechos. Este conocimiento combinado con información histórica que identifique qué clientes han renovado un servicio, cuáles no y qué incentivos fueron ofrecidos a ambos grupos, se puede construir un modelo predictivo que revele qué clientes no renovarán el servicio y ofrecer recomendaciones de los mejores incentivos a ofrecerles.

A continuación se presentan dos casos prácticos del uso de las técnicas de minería de datos en las industrias de servicios financieros y aseguradora. El uso de las técnicas en la mercadotecnia, se desarrollará en el capítulo IV.

II.8.1 Minería de Datos en los Servicios Financieros

Hoy en día, la industria de servicios financieros, comprende una amplia variedad de entidades comerciales como son los bancos, casas de corretaje, bancos mercantiles, compañías inversionistas, compañías financieras, tarjetas de crédito y un sinnúmero de proveedores de servicios que de comercio electrónico vía internet. Típicamente, estas entidades, generan sus ganancias cobrando cuotas pequeñas a un gran volumen de transacciones, en un mercado ferozmente competitivo.

Existen diversas aplicaciones que proveen soluciones para la industria de los servicios financieros: las estadísticas como SAS y SPSS, CSI que es una aplicación que detecta fraude entre transferencias internacionales de fondos vía electrónica, etc.

Pero el énfasis actual de la industria, se centra en la adquisición y retención de los clientes. Por ello, es indispensable determinar a qué clientes dirigir las campañas y predecir cuáles clientes son más propensos a mover sus cuentas hacia un banco u otro.

Las estrategias CRM (*Customer Relationship Management*) son las indicadas para estos casos en que se desea acceder y analizar datos así como generar reportes detallados sobre el comportamiento de los clientes. Como resultado, se pueden dirigir comunicados y publicidad a diferentes tipos de clientes, al conocer sus necesidades específicas o incluso, sucesos de vida importantes para ellos. Es por esto, que la comunicación con los clientes se vuelve indispensable en estos casos.

II.8.2 Minería de Datos en las Aseguradoras

Más allá de las preocupaciones y cálculos obvios que competen el campo actuarial, uno de los mayores retos de esta industria es la detección oportuna del fraude y las compañías aseguradoras desean estar siempre protegidas contra cualquier tipo de pérdida económica. Los tipos de fraude pueden ir desde fraude en el momento del reclamo, evitar el pago de primas, y fraude en la indemnización.

El fraude en el momento del reclamo puede ocurrir en cualquier tipo de seguro. El fraude médico ocurre cuando los proveedores de servicios médicos, proveen tratamientos y servicios innecesarios o llenan reclamaciones para tratamientos y servicios que no fueron llevados a cabo. De igual manera ocurre en los casos en que se ingresan reclamos por reparaciones que no fueron realizadas o de pólizas de vida de personas que ni siquiera han fallecido.

II.9 LA MINERÍA DE DATOS Y LA INTELIGENCIA DE NEGOCIOS

La Inteligencia de Negocios está en constante crecimiento, tanto en términos de capacidad como en la extensión de su implementación. La razón se debe a que la Inteligencia de Negocios y más específicamente, la minería de datos, provee percepciones clave que pueden ser utilizadas para:

- Mejorar la posición en cuanto a la competencia;
- Mejorar los productos y servicios;
- Enfocar los procesos de negocio en los clientes de alto valor;
- Resolver problemas difíciles; y sobre todo
- Mejorar las relaciones con los clientes.

Las principales ventajas competitivas para todas las empresas que hoy en día quieran sobresalir, se originan en la aplicación de la Inteligencia de Negocios en las relaciones con los clientes. La minería de datos es la tecnología clave para proveer tales ventajas competitivas. Junto con los beneficios en ofrecer acceso inmediato a una enorme base mundial de prospectos, la Web ofrece competencia ilimitada y facilidad de compras comparativas para los clientes. Por lo tanto, la satisfacción del cliente debe ser el objetivo. Serán sobresalientes aquellos que ofrezcan una experiencia rica, interactiva y dirigida a los clientes Web y los prospectos.

La Minería de Datos, en particular provee claramente, las mejores capacidades para lograr retornos de inversión potenciales, antes incluso de comenzar un esfuerzo de desarrollo sustancial de un almacén o repositorio de datos. Es por ello que una visión de Inteligencia de Negocios debe estar presente siempre en la empresa, tomando en cuenta las capacidades que puede ofrecer la minería de datos:

- Elementos de datos bien definidos
- Habilidad en el tema al participar en un proyecto
- Datos y resultados históricos suficientes
- Clima organizacional que soporte el tomar riesgos
- Un problema bien definido
- Equipo altamente capaz
- Herramientas capaces para el desarrollo del proyecto

Con el objetivo de entender los potenciales impactos al negocio de un modelo predictivo, una organización debe primero evaluar el modelo con métricas técnicas y después interpretar estas métricas en términos de su impacto financiero en el negocio. Para ello, se tienen que transferir los algoritmos de minería de datos desarrollados por la comunidad científica al área de los negocios, y es entonces cuando la gente de negocios puede evaluar los modelos predictivos desarrollados con las métricas técnicas previas.

Como base para las métricas técnicas se encuentra la Tabla de Confusión que consiste en dos filas (hechos) y dos columnas (predicciones) que comparan resultados verdaderos contra las predicciones de los mismos, como se muestra a continuación:

	Predicción Negativa	Predicción Positiva
Casos Negativos	Verdaderos Negativos TN	Falsos Positivos Error Tipo I FP
Casos Positivos	Falsos Negativos Error Tipo II Falsas Alarmas Riesgos de Manufactura FN	Verdaderos Positivos TP

Fig. II.7 Tabla de Confusión

Para crear una tabla como ésta, se deben establecer los resultados verdaderos y las predicciones de los resultados. Esto se puede llevar a cabo, evaluando cada caso manualmente cuando las observaciones no sean muy grandes y siempre poniendo un límite de tiempo.

Entonces, la tabla de confusión provee información de un modelo predictivo específico en un contexto de negocio también específico, y una vez que los resultados verdaderos para cada caso son conocidos se tienen que ingresar en la tabla de confusión.

	Predicción Negativa	Predicción Positiva
Casos Negativos	9,700	150
Casos Positivos	50	100

Fig. II.8 Ejemplo de Tabla de Confusión Aplicada

La anterior tabla (Fig. II.8) es un ejemplo aplicado a la prevención de fraudes en las aseguradoras. Está derivada de un modelo predictivo elaborado para predecir reclamos fraudulentos en la compañía. Se consideraron 10,000 reclamaciones. Los resultados reales de cada caso se determinaron a través de una investigación manual realizada por expertos en fraudes, y las predicciones fueron determinadas a través del modelo predictivo. De los 10,000 casos estudiados, 9,850 casos no fueron fraudulentos mientras que 150 casos se determinaron como fraudulentos. El modelo predictivo, predijo como “no fraudulento” un número de 9,750 y “fraudulento” un número de 250 casos. Después de comparar los hechos y las predicciones para cada caso, se tiene que hubo 9,700 casos negativos que el modelo predijo correctamente como “no fraudulento”. Hubo también 100 casos que el modelo predijo correctamente como “fraudulento”. Por otro lado, el modelo predijo incorrectamente como “fraudulento” 150 casos falsos positivos y “no fraudulento” para 50 casos falsos negativos.

Una organización puede aprender sobre la calidad de un modelo y el impacto en el negocio, dando seguimiento al número de casos de cada tipo. Los negocios son principalmente afectados por los resultados positivos para cada caso y el común denominador para las métricas de impacto en el negocio son el número de casos positivos y negativos. Por el contrario, la calidad del modelo puede ser caracterizada mediante el análisis de la corrección de las predicciones del modelo que tienen como común denominador el número de predicciones positivas y negativas.

Dado lo anterior, se puede dividir las métricas técnicas como sigue:

	Fórmula	Descripción	Uso en el Negocio
Precisión	$\frac{TN + TP}{TN + FP + FN + TP}$	Tasa de predicciones correctas	(Evitar)
Exactitud	$\frac{TP}{FP + TP}$	Mide la calidad de la predicción: tasa de las predicciones positivas correctas	Eficiencia
Aviso o recall	$\frac{TP}{FN + TP}$	Mide el impacto en el negocio: tasa de los casos positivos predichos correctamente	Efectividad
Especificidad	$\frac{TN}{TN + FP}$	Mide el impacto en el negocio: tasa de los casos negativos predichos correctamente	Economía
Exposición	$\frac{FN}{TN + FN}$	Mide la calidad de la predicción: tasa de las predicciones negativas incorrectas	Error
Selectividad	$\frac{FP + TP}{TN + FP + FN + TP}$	Tasa de las predicciones positivas	Esfuerzo
Tasa de Falsa Alarma	$\frac{FP}{FP + TP}$	Mide la calidad de la predicción: tasa de las predicciones positivas incorrectas	Desperdicio
Tasa <i>Missed Alarm</i> o de Riesgo	$\frac{FN}{FN + TP}$	Mide el impacto en el negocio: tasa de los casos negativos predichos incorrectamente	Riesgo

Fig. II.9 División de las Métricas Técnicas

La *exactitud* mide la tasa de predicciones correctas para las predicciones positivas, es decir, la eficiencia de las predicciones positivas; mientras que la *exposición* mide la tasa de predicciones incorrectas sobre todas las predicciones negativas, es decir, la tasa de error entre las predicciones negativas. Estas dos medidas, nos indican la calidad del modelo. La *especificidad* mide la tasa de predicciones correctas para casos negativos, esto es, el impacto en el negocio

en términos de economía en tanto la clasificación los casos negativos que no requieren de ninguna acción; mientras que el *aviso o recall* mide la tasa de predicciones correctas para casos positivos, esto es, la efectividad de detectar los casos positivos. Estas dos medidas indican el impacto en el negocio.

También se cuenta con medidas que indican las características generales del modelo, relacionando los datos a través de las filas y columnas de la tabla. La *selectividad* evalúa la tendencia de un modelo para realizar predicciones positivas y *la precisión* mide la tasa de predicciones correctas a través de todos los casos.

La tasa de *falsa alarma* mide el desperdicio, lo opuesto a la eficiencia y la tasa *missed-alarm*, mide el riesgo, es decir, lo opuesto a la efectividad. Finalmente, la selectividad relaciona el esfuerzo requerido para actuar sobre predicciones positivas.

A pesar de que todas estas métricas técnicas pueden proveer vistas interesantes sobre la calidad de los modelos y el impacto en los negocios, no pueden contestar la más fundamental de las preguntas que se formulan en un negocio sobre los modelos predictivos: ¿puede un modelo predictivo beneficiar a la organización y cuál de los modelos provee mayores beneficios?

El impacto de los modelos predictivos está dado por una serie de factores distintos. El primer factor financiero es el llamado *costo por evento C(E)*, si se predice un evento adverso como el fraude, o el *beneficio por evento* si se predice un evento benéfico como la respuesta a una campaña de mercadotecnia. El segundo factor a considerar es el *riesgo* o la *probabilidad p*, de que ocurra un evento predicho. El riesgo entonces puede ser utilizado para calcular el *costo esperado*, resultado del producto entre el riesgo y el costo por evento. El tercer factor es el *costo de la acción C(A)* y es el factor de llevar a cabo una determinada acción. Por último, una segunda probabilidad tiene que considerarse, y es la probabilidad de que la acción nos llevará al resultado deseado. Esta probabilidad caracteriza la *efectividad* de la acción y se denota por *E*. El producto de multiplicar el costo esperado por la efectividad de tomar una acción es el *beneficio esperado* de tomar dicha acción.

Derivado de lo anterior, resulta una fórmula para calcular el impacto financiero esperado resultante de tomar una acción para un caso individual con riesgo p:

$$E(I) = C(E) \times p \times E - C(A)$$

A un nivel de conjuntos de datos completo, el impacto en el negocio de un modelo *I(M)*, es el costo que se espera evitar menos el costo de llevar a cabo una acción a través de todos los casos. La organización puede esperar evitar el costo de un evento adverso para una fracción de todas las predicciones positivas, y esta fracción es cuantificada por la efectividad *E* del procedimiento de prevención que la organización implemente. El costo de llevar a cabo una acción, por otro lado incurre en cada una de las predicciones positivas.

La fórmula resultante del impacto financiero esperado, I , del modelo predictivo M es entonces:

$$I(M) = TP \times C(E) \times E - (FP + TP) \times C(A)$$

En conclusión, las métricas técnicas no pueden utilizarse por sí solas para determinar si un modelo predictivo determinado puede beneficiar o dañar a la organización y los costos deben ser tomados en cuenta para deducir los beneficios en el negocio.

III. TOMA DE DECISIONES

III.1 LAS DECISIONES Y LOS ENCARGADOS DE TOMARLAS

Decisión es la elección entre alternativas múltiples, usualmente hecha con un justo rango de raciocinio. En un ambiente organizacional que se jacte de tener un sistema de Inteligencia de Negocios desarrollado, las decisiones afectan el desarrollo de un plan estratégico y como consecuencia, implican decisiones de inversión sustanciales, la definición de iniciativas de mercadotecnia y sus predicciones de ventas relacionadas así como el diseño de un plan de producción que permita a los empleados y los recursos tecnológicos disponibles a ser empleados de una manera efectiva y eficiente.

El proceso de toma de decisiones forma parte de una estrategia de resolución de problemas y se refiere al proceso mediante el cual un grupo de individuos trata de acortar la distancia que existe entre las condiciones de operación actuales del sistema (*as is*) y las condiciones mejoradas supuestas, que se desean alcanzar en un futuro (*to be*).

Para entender los problemas que se enfrentan al tomar una decisión se debe tomar en cuenta el número de factores involucrados en él: uno, dos, cuatro o más.

- Un Factor Involucrado: el encargado de la toma de decisiones se enfoca en un aspecto muy específico o limitado del problema, el objetivo es entonces, exclusivamente el resultado. La solución derivada de este tipo de decisiones suele ser muy pobre. Por ejemplo, el recortar costos ya que esto reduce la capacidad y puede incrementar las pérdidas. Otras aplicaciones típicas son los costos, las ganancias o el crecimiento.
- Dos Factores Involucrados: Compara la relación de causa-efecto que existe en muchas situaciones. Ejemplos típicos son el análisis costo-beneficio y los trade-offs.
- Cuatro Factores Involucrados: Es una extensión de los ejemplos de Dos Factores y se utiliza cuando el problema involucra el equilibrio a nivel organizacional en los que se deben tener en cuenta tanto las fuerzas internas de la organización como las fuerzas externas que afectan el desempeño de la misma, por ejemplo el tener como meta responder a las demandas de los clientes sin sacrificar las metas internas de la organización. Una aplicación son los trade-offs duales.
- Múltiples Factores Involucrados: Es en este caso en el que se consideran todos los factores que afectan una situación determinada, desde el flujo de información, los procesos tecnológicos y el comportamiento de los individuos. Es ideal para un gerente o directivo de la organización saber lidiar con este tipo de casos, para poder discernir entre cuáles factores son relevantes en una situación dada y con los que se pueda ser capaz de resolver un problema.

Por lo anterior, para que un proceso de resolución de problemas pueda llevarse a cabo de una manera óptima, es recomendable tomar en cuenta diversos factores. El diagrama III.1 ilustra la estructura de un proceso de resolución de problemas:

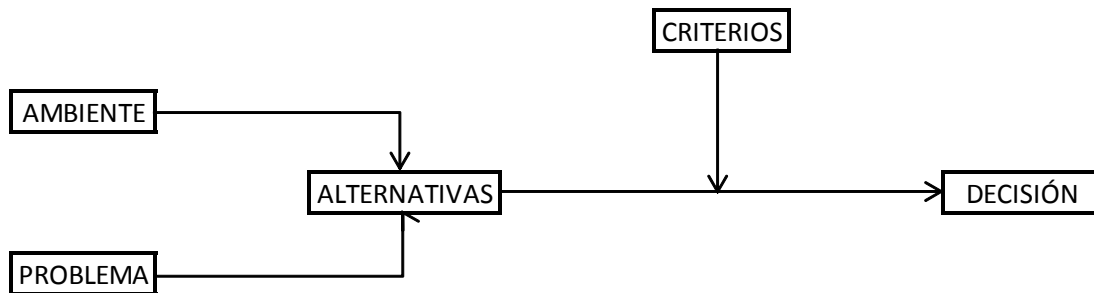


Fig. III.1 Estructura de un proceso de resolución de problemas

Las *alternativas* representan las posibles acciones encaminadas a la resolución de un problema dado y que ayudan a alcanzar el objetivo planeado. Puede haber desde dos hasta un número infinito de alternativas para elegir, en cuyo caso se deben analizar y probar las ventajas y desventajas de cada una de ellas. Los *criterios* son las medidas de efectividad de las diversas alternativas y corresponden a los diferentes tipos de operación del sistema. La toma de una decisión correcta implicará la selección de la alternativa que cumpla el mejor criterio de operación entre todas las alternativas existentes.

El proceso de evaluación de alternativas puede ser dividido en dos etapas:

- **Exclusión:** es la etapa en que las reglas de compatibilidad y restricciones son aplicadas a las acciones alternativas que se identificaron de origen. Es en esta etapa en la que se descartarán algunas alternativas.
- **Evaluación:** Las alternativas viables son comparadas entre sí de acuerdo a los criterios de desempeño y operación para identificar la mejor decisión como la oportunidad ideal.

La figura III.2 muestra cómo se dividen las alternativas y cómo, eligiendo entre diversos criterios se puede llegar a la toma de una decisión correcta:

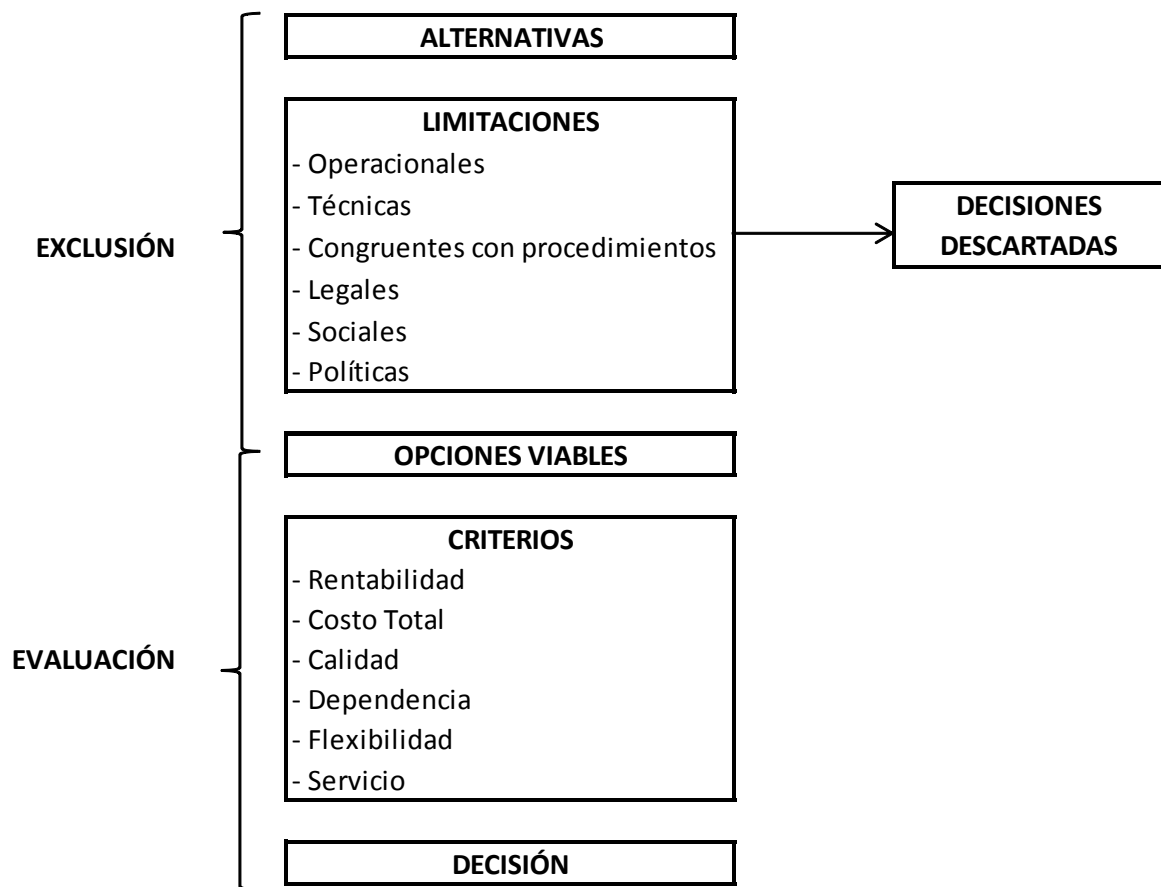


Fig. III.2 División de las alternativas para la toma de decisiones

Por otro lado, los diversos *critérios* dentro del proceso de toma de decisiones que se pueden tomar en cuenta son:

- **Criterios económicos:** Son los que ejercen mayor influencia en este tipo de procesos y son frecuentemente dirigidos a esfuerzos de minimización de costos y maximización de ganancias.
- **Criterios técnicos:** Se deben descartar las opciones que no sean técnicamente factibles.
- **Criterios legales:** Antes de adoptar cualquier decisión tomada, se debe verificar que sea compatible con el marco legal en el que opere la organización.
- **Criterios éticos:** La decisión debe regirse bajo los principios éticos y sociales de la comunidad a la cual pertenezca el sistema
- **Criterios congruentes con los procedimientos:** Una decisión puede caber dentro de los criterios económicos, legales y sociales pero puede no ser viable debido a limitaciones culturales dentro de la organización que tienen que ver con los procesos y prácticas ya establecidas.
- **Criterios políticos:** Se deben tomar en cuenta las consecuencias políticas de una decisión específica.

Los encargados de la toma de decisiones son elementos clave para asegurar que los resultados deseados se alcancen. Sin embargo, el alcanzar estos resultados requiere también el involucramiento de toda la organización para que las decisiones sean implementadas correctamente. Los encargados de la toma de decisiones primero tienen que identificar las metas y los objetivos y en este punto se puede involucrar a otros elementos de la organización; también se debe contar con los recursos y planear las acciones necesarias para alcanzar los objetivos deseados.

En muchas ocasiones, las decisiones tomadas originarán confrontaciones entre los miembros de la organización por lo que se requiere un encargado de toma de decisiones bien enfocado y que se concentre en acciones rápidas y resultados; es por eso que la aceptación de toda la organización es importante.

El encargado de la toma de decisiones se ve afectado por varios factores para desempeñar sus funciones: el ambiente en el que se encuentra la organización, es decir, todos los factores fuera de la empresa que impactan de mayor o menor manera a la organización (clientes, proveedores, bancos, gobierno, etc.); la propia organización y todas las relaciones de trabajo que se tejen al trabajar con otras personas; las demandas para la realización de las tareas, es decir, las habilidades, experiencia, conocimiento y competencias técnicas con que se debe contar para operar de una manera efectiva; y las necesidades personales, ya que como seres humanos, constantemente nos vemos afectados por nuestras emociones, personalidad, valores que afectarán, de igual manera, las decisiones tomadas.

Los gerentes o directivos de una organización, pueden ayudarse de las herramientas de análisis tanto de Inteligencia de Negocios como de Minería de Datos para manejar efectivamente los problemas a los que se enfrentan.

III. 2 LOS TIPOS Y ESTILOS DE DECISIONES

Existen varios tipos de decisiones que pueden clasificarse de acuerdo a su naturaleza y su objetivo.

De acuerdo a su *naturaleza* las decisiones pueden clasificarse como sigue:

- **Decisiones estructuradas:** Están basadas en un procedimiento de toma de decisiones bien definido y recurrente. En la mayoría de los casos, las decisiones estructuradas pueden ser originadas de acuerdo a un algoritmo más o menos explícito para los encargados de la toma de decisiones y por lo tanto, ser más adaptables para su automatización.
- **Decisiones no estructuradas:** En este tipo de decisiones, los datos fuente, los resultados o bien, los procesos de transformación no pueden ser descritos a detalle. Este tipo de eventos suelen ocurrir cuando nos enfrentamos a un proceso de toma de decisiones por primera vez. En este tipo de decisiones, es fundamental la participación de elementos clave de la organización que posean el conocimiento preciso así como la utilización de sistemas de Inteligencia de Negocios que provean el soporte necesario en cuanto al acceso a la información necesaria para la toma de decisiones.
- **Decisiones semi-estructuradas:** Se da cuando algunas fases son estructuradas y otras no. La mayoría de las decisiones a las que se enfrenta una organización son de este tipo. Para las fases no estructuradas el apoyo de los sistemas de Inteligencia de Negocios es vital ya que se traduce en el acceso a la información de manera oportuna y versátil. Para las fases estructuradas, es posible contar con un soporte activo a través de modelos matemáticos y algoritmos que nos permitan automatizar partes significantes del procedimiento.

De acuerdo a su *objetivo* las decisiones se clasifican en:

- **Estratégicas:** Afectan a la organización entera o a buena parte de ella, modifican los objetivos y políticas de una empresa y son tomadas por altos niveles de la organización
- **Tácticas:** Afectan sólo a ciertas partes de la empresa o exclusivamente a un departamento de la misma.
- **Operativas:** Se refieren a actividades específicas dentro de la organización y tienen el impacto más modesto en el futuro.

El cuadro III.3 presenta las características de la información necesaria para la toma de decisiones de acuerdo a su objetivo:

	Operativas	Tácticas	Estratégicas
Precisión	Alta	↔	Baja
Nivel de detalle	Detallada	↔	Colectivo
Horizonte en el tiempo	Presente	↔	Futuro
Frecuencia de uso	Alta	↔	Baja
Fuente	Interna	↔	Externa
Objetivo de información	Cuantitativo	↔	Cualitativo
Naturaleza de la información	Limitada	↔	Amplia
Antigüedad de la información	Presente	↔	Pasada

Fig. III.3 Características de la información

Las decisiones también pueden ser vistas desde el punto de vista personal y psicológico y el nivel de interacción y comunicación que requieren:

- **Decisiones rutinarias:** Se consideran decisiones programadas y están encaminadas a alcanzar las metas de la organización siguiendo políticas y reglas preestablecidas.
- **Decisiones creativas:** Son decisiones no programadas y nacen de la necesidad de tener nuevas alternativas para el manejo de problemas más complejos.
- **Decisiones negociadas:** En ellas se debe involucrar a los participantes en varios niveles de la organización ya que se tienen que resolver los conflictos entre las metas y los métodos para manejarlas.

Existen, por otro lado, diversos estilos para la toma de decisiones que pueden estar basados tanto en los valores como en la percepción de quienes las toman. El estilo de toma de decisiones describe la manera en que un gerente o directivo toma las decisiones y depende de varios factores como son el contexto en que la decisión es tomada, la manera en que el encargado de tomar las decisiones percibe y entiende información clave, así como de sus valores y juicios.

Los estilos para tomar decisiones pueden dividirse con base en los factores que afectan la toma de una decisión así como en la situación bajo la que se lleva a cabo la misma. Estos estilos son:

- **Directivo:** Este individuo cuenta con poca tolerancia a la ambigüedad y baja complejidad cognitiva. Se enfoca en decisiones técnicas.
- **Analítico:** Tiene la habilidad para hacer frente a situaciones novedosas. Disfruta la resolución de problemas y lucha por alcanzar el máximo en una situación dada.
- **Conceptual:** Cuenta con complejidad cognitiva y orientación hacia las personas, utiliza datos de diversas fuentes y considera varias alternativas.
- **Del comportamiento:** A pesar de su baja complejidad cognitiva, este gerente o directivo tiene una profunda preocupación por la organización y el desarrollo de la gente.

El cuadro III.4 ilustra a detalle los cuatro estilos en los que se puede dividir a los encargados de tomar una decisión:

	Lógico	Relacional	
Tolerancia a la ambigüedad	<p style="text-align: center;">ANALÍTICO</p> <p>Gusto por la resolución de problemas Desea la mejor respuesta Desea tener el control Usa un número de datos considerable Gusto por la variedad Gusto por innovar Utiliza análisis puntual NECESITA DE RETOS</p>	<p style="text-align: center;">CONCEPTUAL</p> <p>Orientado a objetivos Con una visión amplia Creativo Humanista/artístico Inicia nuevas ideas Visión hacia el futuro ES INDEPENDIENTE Y DESEA RECONOCIMIENTO</p>	Reflexivo (ideas)
Necesidad de estructura	<p style="text-align: center;">DIRECTIVO</p> <p>Espera resultados Agresivo Actúa con rapidez Utiliza reglas Utiliza su intuición Verbal NECESITA PODER</p>	<p style="text-align: center;">DEL COMPORTAMIENTO</p> <p>Proporciona soporte Persuasivo Empático Se comunica fácilmente Prefiere las reuniones Utiliza un número de datos limitado NECESITA LEGITIMACIÓN</p>	Activo (acciones)
	Técnico	Social	

Fig. III.4 Estilos en la toma de decisiones

III.3 EL PROCESO Y LOS SISTEMAS DE SOPORTE DE LA TOMA DE DECISIONES

La naturaleza de un proceso de toma de decisiones depende de varios factores como son:

- Las características de la organización donde se lleve a cabo el proceso
- Las actitudes subjetivas de los encargados de la toma de decisiones
- La disponibilidad de metodologías apropiadas para la resolución de problemas
- La disponibilidad de herramientas efectivas que soporten las decisiones

El proceso de toma de decisiones se divide en 5 fases que son: inteligencia, diseño, elección, implementación y control.

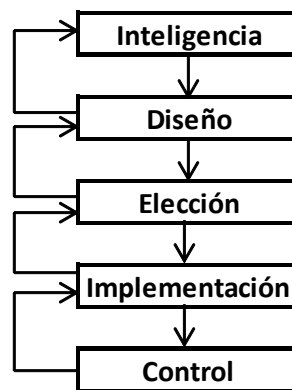


Fig. III.5 El proceso de toma de decisiones

- **Inteligencia:** En esta fase, las tareas del encargado de la toma de decisiones son: identificar, circunscribir y definir explícitamente el problema que se desea estudiar y resolver. El análisis del contexto y de toda la información disponible llevan a los encargados de la toma de decisiones a detectar las señales y síntomas que necesitan de una acción correctiva para mejorar el desempeño del sistema.
- **Diseño:** Se desarrollan y planean las acciones encaminadas a la resolución del problema. Es en esta fase en donde se generan las posibles soluciones que ayudarán a alcanzar el objetivo propuesto, utilizando la experiencia y creatividad de los encargados de la toma de decisiones.
- **Elección:** Una vez que las alternativas han sido identificadas, es necesario la evaluación de las mismas con base en criterios de desempeño que sean significativos. Es aquí donde los modelos matemáticos y la resolución de los mismos juegan un papel importante.
- **Implementación:** Cuando se ha elegido la mejor alternativa, se debe transformar en acciones mediante un plan de implementación, lo que significa la asignación de responsabilidades a aquellos involucrados en el plan de acción.
- **Control:** Finalmente, y una vez que las acciones fueron implementadas, es necesario verificar y revisar que las expectativas originales fueron satisfechas y que los resultados de las acciones concuerdan con las intenciones originales. Se deben

detectar y medir las diferencias entre los indicadores de desempeño identificados en la fase de elección y los que se observen al final de la implementación.

Los aspectos más relevantes que caracterizan un proceso de toma de decisiones pueden ser resumidos en los siguientes puntos:

- Las decisiones son tomadas por un grupo de personas y no por un individuo en particular.
- El número de alternativas de acción puede ser muy alto y, algunas veces, ilimitado.
- Los efectos de una decisión dada usualmente aparecen tiempo después de haber sido tomada y no inmediatamente.
- Las decisiones tomadas en una organización, ya sea pública o privada, normalmente están interconectadas y tienen efectos colaterales. Cada decisión tiene consecuencias para muchos individuos en varias partes de la organización.
- Durante el proceso de toma de decisiones, se necesita que usuarios clave, accedan a los datos y a la información y que trabajen en ellos basándose en un marco conceptual y de análisis.
- En la mayoría de los casos, el proceso de toma de decisiones tiene varias metas, con indicadores de desempeño distintos que, posiblemente, entren en conflicto entre sí.
- Muchas decisiones son tomadas en un contexto confuso y conllevan factores de riesgo.
- Los experimentos que se basan en un método de prueba y error pueden ser de alto riesgo y muy costosos como para ser considerados una práctica dentro del proceso de toma de decisiones.
- El dinamismo con que una empresa opera, afectado por la presión de un mercado competitivo, implica que los usuarios clave dirijan sus esfuerzos a la toma de decisiones de una manera rápida y dentro de un marco de tiempo reducido.

Para ayudar a que la toma de decisiones cumpla con los objetivos deseados, es necesario que se haga de manera efectiva, que se apoye de modelos matemáticos tomados de disciplinas como la investigación de operaciones o la estadística y que sea integral, es decir que involucren a los individuos de varias áreas de la organización; es por esto que lo ideal es que un Sistema de Apoyo a las Decisiones (**DSS** por sus siglas en inglés) sea implementado en la empresa. Los sistemas de apoyo a las decisiones están conformados por varios módulos:

- **Manejo de datos:** El módulo de manejo de datos incluye una base de datos diseñada para contener los datos requeridos en los procesos de toma de decisiones al cual está direccionado el sistema de apoyo a las decisiones. En la mayoría de las aplicaciones, esta base de datos es un *data mart* que normalmente se conecta con el almacén de datos de la compañía y es el repositorio principal de los datos disponibles para desarrollar los análisis de inteligencia de negocios.
- **Manejo de modelos:** Este módulo provee a los usuarios finales, un conjunto de modelos matemáticos derivados de la investigación de operaciones, estadísticas y análisis financieros. Usualmente son modelos relativamente simples que permiten llevar a cabo las investigaciones analíticas necesarias para el proceso de toma de decisiones. En algunas aplicaciones, este módulo se puede integrar por modelos

complejos llamados *modelos externos*, creados para llevar a cabo tareas de análisis específicas, por ejemplo, un modelo de optimización a gran escala formulado para desarrollar el plan logístico anual de una empresa manufacturera.

- **Interacciones:** En la mayoría de las aplicaciones, los usuarios responsables del conocimiento, utilizan el sistema de apoyo a las decisiones de manera interactiva para llevar a cabo sus análisis. El módulo responsable de estas interacciones está diseñado para ser alimentado con datos proporcionados por usuarios de la manera más sencilla e intuitiva, usualmente a través de una interfaz gráfica en un dispositivo web, para después devolver la información extraída y el conocimiento generado por el sistema de una forma gráfica apropiada.
- **Manejo del conocimiento:** El módulo de manejo del conocimiento del sistema de apoyo a las decisiones, también está interconectado con el sistema integral de manejo del conocimiento de la compañía. Esto permite, a los encargados de la toma de decisiones, recurrir a las distintas formas de conocimiento colectivo que usualmente no tienen una estructura definida y que representan la cultura corporativa de la organización.

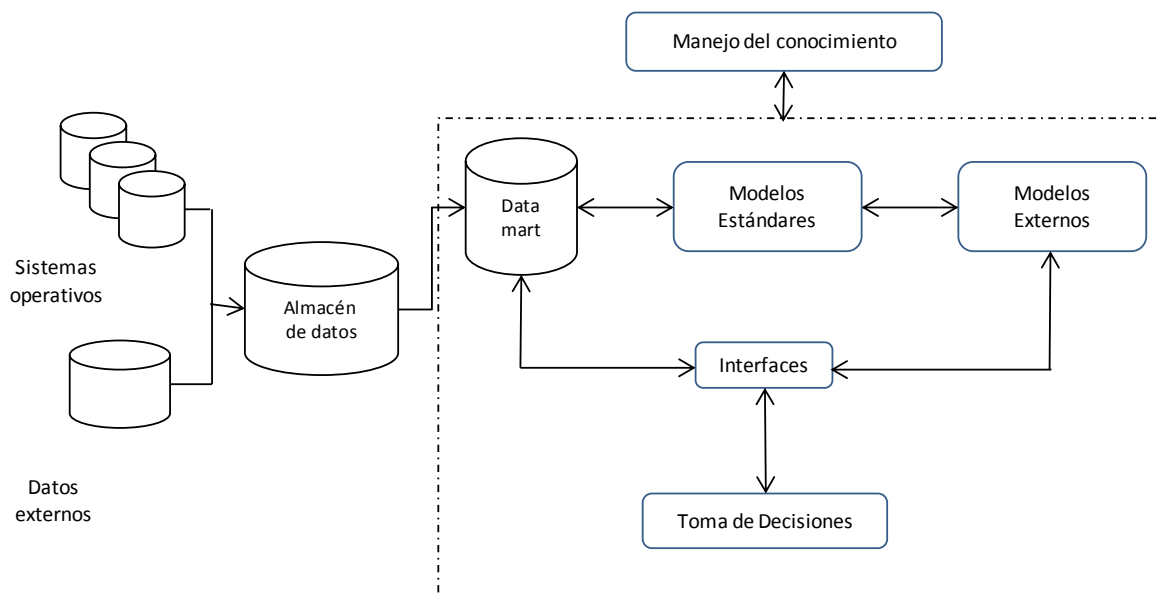


Fig. III.6 Estructura de un Sistema de Apoyo a las Decisiones (DSS)

III.4 LA TOMA DE DECISIONES DENTRO DE LAS ORGANIZACIONES

La toma de decisiones dentro de las organizaciones ha sido objeto de estudio desde principios de los años 70 y normalmente ha estado basada en el comportamiento particular de los individuos dentro ellas. Se considera a los individuos como parte fundamental de los procesos y se toma en cuenta su actitud y desempeño frente a los diferentes factores, internos y externos, que afectan el equilibrio organizacional, así como las relaciones entre ellos.

Autores como Shull, Delbecq y Cumming identificaron cinco factores dentro de las organizaciones que afectan los tipos de decisiones y la manera en que éstas son tomadas:

- Estructura: Describe la relación entre los miembros de una organización
- Roles: Define el comportamiento necesario para facilitar las tareas
- Procesos: Describe la manera de proceder hacia el cumplimiento de objetivos
- Estilo: Relaciona las interacciones interpersonales con el estrés, afinidad y las consecuencias del éxito y el fracaso
- Normas: El marco de referencia compartido, las presiones sociales y las prescripciones con respecto al comportamiento, creencias, o sentimientos que son reforzados por medio de sanciones.

Utilizando estos cinco factores, describen también tres tipos de decisiones básicas que son: decisiones rutinarias, creativas y negociadas.

No se puede considerar que en una organización, la toma de decisiones se haga de manera unilateral o que sólo existan ciertas personas o grupos encargados de llevarlas a cabo sino que se trata de procesos en los que lo ideal es que sean tomados en cuenta e intervengan individuos pertenecientes a distintos niveles organizacionales. El reconocimiento de las necesidades individuales de manera explícita es un punto clave que determina la efectividad de las decisiones en la organización. Este tipo de decisiones están directamente relacionadas con la pregunta de qué necesita un buen gerente o directivo, así como con el hecho de que si se cuenta con colaboradores motivados, éstos harán las cosas por propia iniciativa.

Pero no todas las organizaciones pueden funcionar de la misma manera en el proceso de toma de decisiones, ya que las necesidades, funciones, controles y problemas dependen directamente de la fase de vida en que se encuentre la empresa, es decir, de la antigüedad que tenga la empresa y los colaboradores que en ella participan.

Se han identificado cinco fases por las cuales pasan las organizaciones y que afectan claramente en las funciones de dirección y requerimientos de información.²

Fase	Función	Organización	Control	Crisis
I	Innovación	Informal	Crear	Liderazgo
II	Transición	Funcional	Dirigir	Autonomía
III	Crecimiento	Descentralizada	Delegar	Control
IV	Consolidación	Grupos	Coordinar	<i>Red Tape **</i>
V	Adaptación	Equipos	Colaborar	Sobrevivencia

Fig. III.6 Fases de vida de las organizaciones

Esto quiere decir que cuando recién creada, una organización necesita que se asigne a sus trabajadores funciones de innovación y que sean creativos, lo que llevará indirectamente a la definición de líderes, a pesar de que, la organización de funciones se haga de manera informal. En una segunda etapa, o de transición, es necesario dirigir los esfuerzos (el líder aquí es una pieza clave) y comenzar a separar esfuerzos con el fin de que cada uno de los colaboradores sea autónomo y enfocado a funciones específicas. Posteriormente, en una etapa de crecimiento es necesario comenzar a controlar los procesos y esto no se puede hacer de otra manera que delegando funciones para descentralizarlas. En una etapa de consolidación, la organización se divide por grupos quienes coordinan actividades específicas encaminadas a las prácticas **Red Tape** **. Finalmente, una empresa consolidada, tiene que dirigir sus esfuerzos a mantenerse en pie, a sobrevivir y esto no se logrará de otra manera que con la colaboración mutua de los distintos equipos que se han ido formando y con la inminente adaptación de todos sus integrantes.

Con respecto a las funciones de la organización, las respectivas tareas de gestión y los requerimientos de información las fases de la organización quedan divididas como sigue:

Fase	Función	Gestión	Requerimientos de Información
I	Innovación	Participativa	Reportes de status
II	Transición	Flexible	Reportes por solicitud
III	Crecimiento	Analítica	Automatizada, repores en tiempo real
IV	Consolidación	Que inove	Sistemas de soporte de decisión (DSS)
V	Adaptación	Que coordine	Procesamiento en red

Fig. III.7 Fase, función, gestión y requerimientos de información en las organizaciones

Así que, conforme una organización alcance su madurez, la participación de los gerentes o directivos tiene que ir cambiando y pasar de ser llevada a cabo de una manera participativa, en la que todos los colaboradores estén involucrados en todos los procesos, a llegar al punto en que sólo se coordinen los esfuerzos de dichos colaboradores para alcanzar los objetivos. De igual manera los requerimientos de información tienen que ir mudando ya que al principio sólo se necesitarán reportes del status de las tareas que muy probablemente, pueden ser elaborados por las mismas personas encargadas de la gestión y la dirección; posteriormente ya se pueden ir solicitando a especialistas en el manejo de la información dichos reportes para que, una vez que se cuente con el equipo necesario, estos puedan ser automatizados y así estén al alcance de toda la organización en un marco de tiempo real. Finalmente, serán creados sistemas de apoyo a las decisiones (DSS) descritos en el punto III.3 para finalmente tener una red que proporcione información a toda la organización.

No importa en qué fase de vida se encuentra una organización, existen cinco factores que afectan las situaciones de toma de decisiones en los que se utiliza el poder que una persona tiene dentro de la misma y cómo lo emplea:

*** Se conoce como prácticas Red Tape al conjunto o secuencia de formas y procedimientos requeridos para ganar la aprobación burocrática de algo, especialmente en situaciones complejas y que consumen tiempo. The American Heritage Dictionary of the English Language, 4a Edición, Houghton Mifflin Company*

- Poder: Es el resultado de la intervención directa o indirecta de las preferencias en el proceso de toma de decisiones.³ Por lo tanto, la manera de compartir el poder determina cómo el gerente o directivo quiere ejercer el poder disponible. El poder también se refiere a la manera en que el gerente ejerce su autoridad.
- Autoridad: Es la legitimidad que el gerente tiene para los actos que están bajo su supervisión.
- Poder informal: Es un conjunto de relaciones que permiten el cumplimiento de tareas conformadas por actividades interdependientes y que requieren el esfuerzo y cooperación mutuos. La negociación y los consensos con actividades que se llevan a cabo bajo un marco informal de poder.
- Influencia: Es el resultado de la intervención directa o indirecta de las preferencias de un gerente en la toma de una decisión.
- Políticas: Un proceso de acuerdo a políticas es el que incrementa la seguridad dentro de la organización, basado en reducir los efectos de factores que contribuyen a la incertidumbre.⁴

El cuadro III.8 muestra la relación entre el poder y la autoridad que un gerente o directivo ejerce sobre sus colaboradores:

Tipos de Autoridad	Grupo Informal	Políticas de Control	Coaliciones negociadoras	Consenso
	Referente Experto	Roles Legítimos	Resolución de problemas	Poder compartido
	Posición Formal	Obediencia	Motivación	Delegar tareas
		Coercivo Autoritario (Manipulador)	Recompensas Sanciones (Inducción)	Persuasivo Influencias (Participación)

Tipos de Poder

Fig. III.8 Relación entre tipos de poder y autoridad

El cuadro anterior no intenta demostrar que algún tipo de combinación, tipo de poder – tipo de autoridad, sea la ideal, sino que se debe observar cada caso en forma particular. Lo que nunca se debe perder de vista, es que cada uno de los colaboradores en una organización son seres humanos, cada uno con formas de pensar y de actuar distintas e independientes y que tienen una percepción de la palabra “poder” muy particular, esto quiere decir que, existen personas a las que un sistema de recompensas y castigos les puede funcionar para lograr que alcancen sus objetivos mientras que, algunos otros, prefieren recibir una influencia de manera persuasiva mediante un proceso más intelectual en los que el consenso y el sentimiento de tener el poder compartido en la toma de decisiones se vuelve básico.

John Kotter mencionó que a medida que las organizaciones se vuelven más complejas, es difícil, si no imposible para los gerentes, alcanzar sus objetivos de manera independiente o a través de la persuasión o la autoridad formal solamente. Necesitan cada vez más del poder

3. *Managerial Decision Making: A Study of Leadership Styles/ Heller, Frank A./Tavistock Publications Ltd./1971*

4. *Management Control and Decision System/ Patz, Alan E. and Rowe, Alan / Wiley / 1977*

para influenciar a otras personas de las cuales dependen. Más aún, los gerentes efectivos tienden a tener éxito al desarrollar cuatro tipos de poder diferentes que utilizan junto con la persuasión y la influencia sobre los otros. Esto lo realizan con madurez, grandes habilidades y sensibilidad a las obligaciones y a los riesgos que involucran.⁵

Las maneras en que Kotter definió que los gerentes exitosos establecen las relaciones de poder son:

- a. Crear el sentimiento de obligación en los otros, lo que el gerente utiliza para influenciarlos.
- b. Construir una reputación como experto en un campo determinado.
- c. Fomentar la identificación consciente de los otros con el gerente y con sus ideas o posturas.

Para definir una correcta intervención del equipo de trabajo en la toma de decisiones, se puede utilizar el siguiente modelo de Heller⁶ para compartir la participación de cada uno de los integrantes:

	Alta	
Experiencia y habilidades de los subordinados	Alta	Consultar a los subordinados
	Baja	Tomar la decisión por sí solo
		Alta
		Importancia de la decisión

Fig. III.9 Modelo de Heller de participación en las decisiones

Lo anterior, nos lleva a la conclusión de que, entre más capacitado se encuentren los integrantes de un equipo de trabajo, las decisiones grupales se pueden tomar en conjunto, lo que generará, invariablemente un ambiente de trabajo, en el que cada uno de los participantes en la toma de las decisiones se sentirá parte importante de un proceso y de los resultados del mismo.

De cualquier manera, siempre se debe tener en mente la implementación de controles, que serán los indicadores que señalarán de manera efectiva, si una decisión fue correctamente tomada e incluso, cuáles son las repercusiones de la participación activa del equipo encargado de tomarla. El ideal, es tener controles que se anticipen a posibles desviaciones en los proyectos, como se hace en el control de calidad estadístico. Es en estos casos, donde los sistemas programados basados en la Minería de Datos son utilizados, ya que se manejan grandes cantidades de datos y modelos de simulación de resultados que ayudarán a determinar si a partir de la modificación de alguno de los componentes del proceso en general, una decisión puede ser exitosa o infructuosa.

5. *Power: How to Get It and How to Use It/ Kotter, Michael / Free Press / 1985*

6. *Managerial Decision Making: A Study of Leadership Styles/ Heller, Frank A./Tavistock Publications Ltd./1971*

El esquema III.10 muestra cómo intervienen los controles en conjunto con la Toma de Decisiones y la Minería de Datos en una organización que basa su operación en la Inteligencia de Negocios:

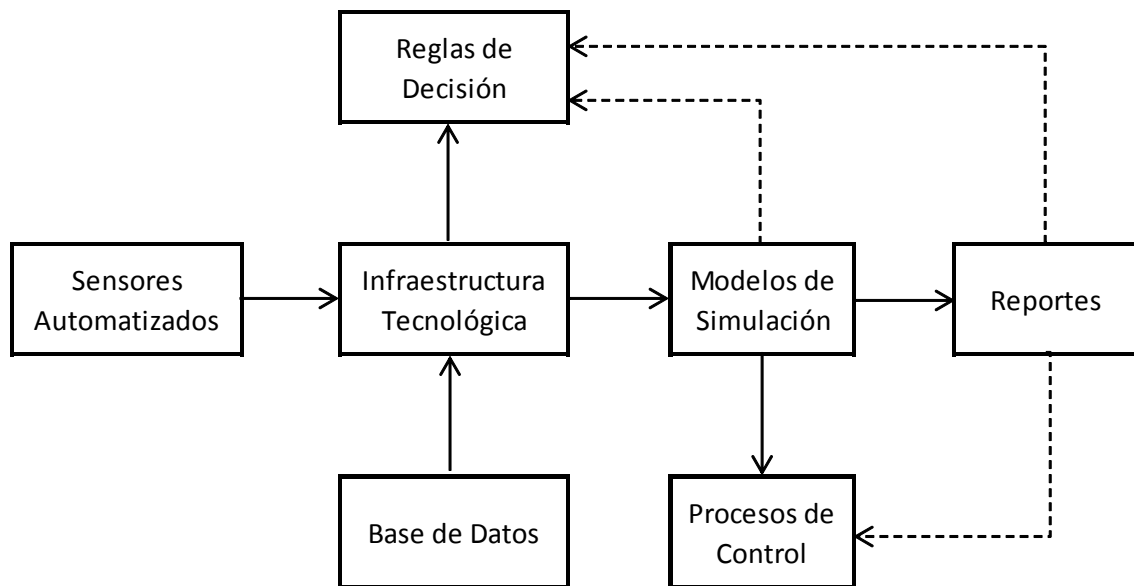


Fig. III.10 Esquema de organización basada en la Inteligencia de Negocios

Los pasos anteriores sumados a la intervención humana en cuanto a conocimientos y experiencia, hacen que se pueda predecir con mayor porcentaje de certeza, los efectos de la toma de decisión.

No se debe dejar de lado el tema de los costos que implica tener procesos de control en la organización. Los costos incrementan exponencialmente cuando se usan sistemas de control más estrictos, mientras que, por otro lado, el valor al que contribuyen los sistemas de control estrictos alcanza su máximo. Se utiliza el valor neto para definir el punto de control óptimo deseado.

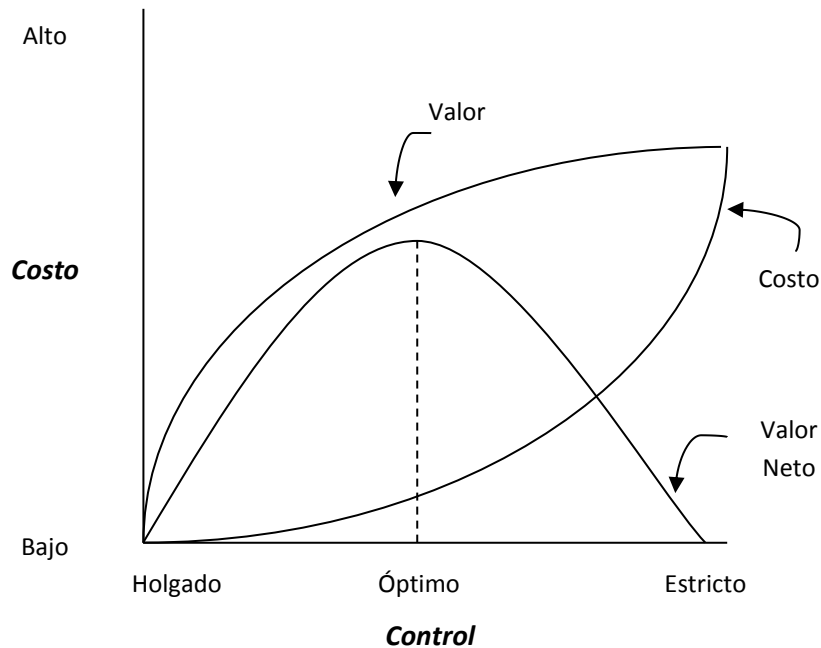


Fig. III.11 Gráfico del Valor Neto

Un entendimiento profundo de la causalidad se necesita para crear controles que faciliten la implementación y que se adapten a condiciones nuevas y cambiantes. La razón por la cual ocurren las variaciones, debe ser determinada para que se tomen las acciones correctivas apropiadas en tiempo y forma. Además, es necesario elaborar un *forecast* o proyección del impacto de las decisiones para que se pueda saber cómo implementar los controles. Como se muestra en la figura III.11, el valor de control alcanza un punto de Rendimientos Decrecientes.

La ley de Rendimientos Decrecientes afirma que cada vez se obtendrá menos producción adicional a medida que se añaden cantidades adicionales de una de las entradas o input, manteniendo el resto de factores constantes. Dicho de otro modo, el producto marginal de cada unidad de una entrada se reducirá a medida que la cantidad de esta entrada aumente, si los otros permanecen constantes; debe entenderse por producto marginal de una entrada la cantidad de producción adicional que se obtiene después de añadir una unidad adicional de este manteniéndose todos los demás constantes.

A pesar de que la efectividad de un sistema de control es típicamente juzgada por qué tan cerca se encuentra el desempeño real de alcanzar un objetivo determinado, otra manera de ver la efectividad es:

$$\text{Efectividad} = \frac{\text{Valor recibido}}{\text{Recursos usados}} \times \text{Costo del sistema de control}$$

Cuando el costo del sistema de control incluye tanto los costos de los recursos como del desempeño y es dependiente del proceso o estilo gerencial utilizado para alcanzar dicho control.

Una herramienta muy útil para establecer los controles de forma preventiva es utilizando la Ley de Pareto o regla del 80-20: el 20% de los individuos o productos son los que causan el 80% de los problemas, lo que lleva también a la conclusión de que, al controlar el 20% de los puntos en un proceso, se puede alcanzar el 80% de éxito deseado. Esta es una base muy poderosa para las medidas de control. La ley de Pareto provee medios directos para dividir cualquier problema en puntos importantes y no importantes. Esta aproximación puede ser utilizada en la toma de decisiones, en el esfuerzo de distribución de una fuerza de ventas, o proyectos de presupuesto. Con respecto a la toma de decisiones, sólo un pequeño porcentaje de las que se toman por un equipo o persona tiene un impacto significativo en la organización. Los ejecutivos efectivos concentran sus esfuerzos en aquellas decisiones que tienen un mayor beneficio. De manera equivocada, en muchos de los casos y organizaciones, se da peso o prioridad equivalente a todas las decisiones que se toman y no se elabora un reconocimiento de su importancia relativa.

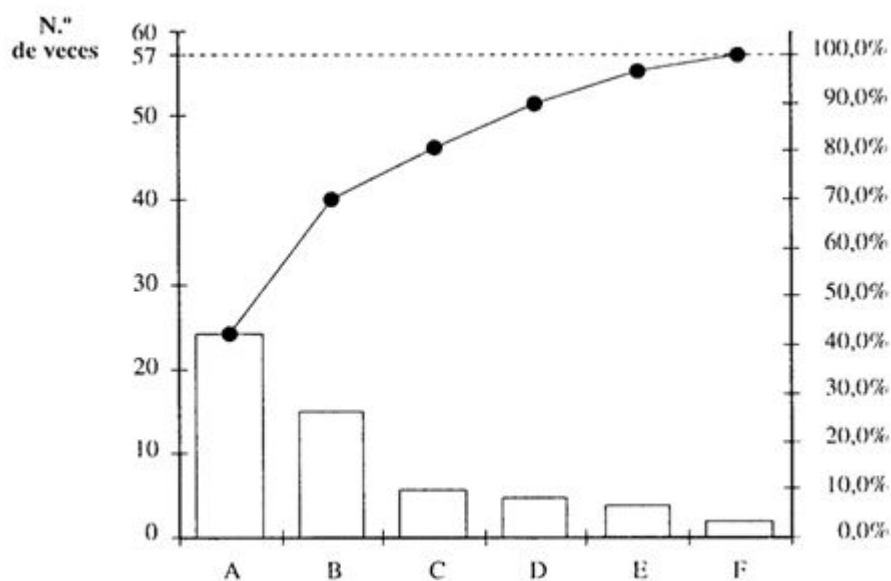


Fig. III.12 Ejemplo de diagrama de Pareto para un caso de reclamaciones

III.5 HERRAMIENTAS Y MODELOS MATEMÁTICOS PARA LA TOMA DE DECISIONES

En este punto se describen algunas herramientas y ejemplos sobre formas estructuradas y analíticas para tener un proceso de toma de decisiones óptimo basado en los tipos de problemas que se desean resolver:

III.5.1 Herramientas Deterministas: Cubre los problemas con un enfoque directo, aplicando herramientas cuantitativas.

III.5.2 Herramientas Complejas: Estos problemas involucran muchas combinaciones por lo que métodos cuantitativos y tecnológicos pueden ser utilizados.

Entre las herramientas **deterministas** se encuentran:

III.5.1.1 Análisis de decisión lineal: En esta técnica, todos los aspectos de un problema son vistos como una serie de sub problemas (distintas maneras de dividir el problema principal en unidades más pequeñas y manejables), cada uno con entradas (datos o información utilizados) y salidas (ya sea la acción tomada o los resultados esperados con respecto al sub problema). En la mayoría de los casos, la salida de uno de los pasos se convierte en la entrada del siguiente, sin embargo esta condición no es requerida para analizar o definir un problema.

III.5.1.2 Tablas de decisión: Esta técnica, requiere que el encargado de la toma de decisiones defina explícitamente todas las condiciones aplicables y qué acciones o decisiones serán tomadas para cada conjunto de condiciones. Se maneja a través de reglas simples y se recomienda utilizarlas sólo para identificar factores relevantes y sus combinaciones como base de decisiones muy claras y precisas.

III.5.1.3 Análisis Costo-Beneficio: Se utiliza en los casos en los que se intenta evaluar el costo de una inversión y es entonces que el gerente o directivo, típicamente balancea los beneficios que se derivarán de dicha inversión. Actualmente el racional de costo-beneficio debe ser menor a uno para que el proyecto pueda contribuir a un mayor valor que su costo.

Entre las herramientas **complejas** se tienen los siguientes ejemplos:

III.5.2.1 Simulación: Se puede utilizar la simulación para tipos de decisiones específicas como son:

- Determinar la capacidad necesaria del equipamiento, instalaciones y labores para enfrentar demandas impredecibles por parte de los clientes.
- Examinar los niveles alternativos de demanda y la capacidad de los sistemas para hacerles frente.

- Examinar la construcción de los inventarios y la utilización del equipo necesarias para hacer frente a los requerimientos de los clientes.
- Desarrollar reglas de decisión programadas para mantener un inventario mínimo mientras se cumple con los requerimientos de entregas

Una consideración importante en la aplicación de la simulación es el modelo usado para describir la gran complejidad que existe en un problema actual. Si el problema ya está siendo observado y puede ser descrito, entonces establecer el nivel de confianza con respecto a qué tan cerca dicha simulación representa realmente al problema.

III.5.2.2 Decisiones con atributos múltiples: Este tipo de toma de decisiones involucra varios factores o atributos que necesitan ser considerados al examinar las alternativas de resolución de un problema. Sin embargo, los problemas requieren de el encargado de la toma de la decisión identifique de forma explícita, las alternativas y criterios que serán utilizados para elegir la opción más deseable.

III.5.2.3 Sistemas de Apoyo a las decisiones: Los data marts, los almacenes de datos, las tablas dimensionales, las tablas de hechos, etc. descritas en el capítulo I componen estos sistemas que soportan la toma de decisiones.

Finalmente están las decisiones que involucran **riesgo**. Evidentemente, cualquier tipo de decisión involucra directa o indirectamente un riesgo y es aquí cuando el equipo encargado de la toma de decisiones puede elegir entre ignorar dicho riesgo o no permitirse tomarlo.

III.5.2.3.1 Reglas de probabilidad: Tomar decisiones con el factor riesgo y la incertidumbre involucrada requiere, al menos, un conocimiento básico de probabilidad y, obviamente, basarse en información histórica almacenada en las bases de datos.

Por ejemplo, en una empresa financiera se tiene que:

PR = bajo riesgo = el porcentaje de todas las personas que solicitan un préstamo que son un riesgo

DL = préstamos quebrantados = el porcentaje de todos los préstamos que no serán pagados

Y se requiere saber:

PRR = el porcentaje de riesgos bajos que pagarán sus préstamos

GCD = el porcentaje de buenos créditos que serán quebrantados

Se buscan entonces los siguientes racionales:

$$PR = \frac{\text{Número de personas calificadas como de bajo riesgo}}{\text{Número total de personas que solicitan un préstamo}}$$

$$DL = \frac{\text{Número de personas que serán quebrantadas}}{\text{Número total de personas a las que se les otorga un préstamo}}$$

$$PRD = \frac{\text{Número de riesgos bajos que serán quebrantados}}{\text{Número total de riesgos bajos}}$$

$$GDC = \frac{\text{Número de buenos créditos con riesgo de quebrantar}}{\text{Número total de buenos créditos con riesgo}}$$

Finalmente, combinando estos factores, se pueden estimar las pérdidas potenciales del banco, una vez que se revisan los registros.

En cuanto a los modelos matemáticos que apoyan la toma de decisiones y como previamente se introdujo en el capítulo I, se incluyen:

- III.5.3 Modelos predictivos:** Como se vio en el punto II.6.4, estos modelos juegan un rol primordial en los sistemas de inteligencia de negocios ya que están ubicados de manera superior con respecto a otros modelos matemáticos y en general, a todo el proceso de toma de decisiones. Las predicciones permiten alimentar información en los procesos de toma de decisiones impactando en las estrategias, la investigación y el desarrollo, la administración y el control, la mercadotecnia, la producción y la logística. Como ejemplo de modelos de explicativos están la regresión y la clasificación y los modelos de series de tiempo.
- III.5.4 Modelos de reconocimiento y aprendizaje de patrones:** Son la base de los modelos predictivos y su objetivo es entender los mecanismos que regulan la habilidad de extracción del conocimiento con base en experiencia anterior que pueda ser aplicado en el futuro. Se aplican en varios campos de toma de decisiones como son el reconocimiento de textos, imágenes y sonidos; diagnósticos médicos, mercadotecnia relacional para la segmentación y elaboración de perfil de los clientes, así como en la detección de fraudes.
- III.5.5 Modelos de optimización:** Dado un problema, el encargado de la toma de decisiones define un conjunto de decisiones factibles y establece los criterios para la evaluación y comparación de las distintas alternativas, como los costos o los pagos. Es en este punto en el que el encargado de la toma de decisiones debe identificar la decisión óptima de acuerdo con el criterio de evaluación definido. En general, los modelos de optimización surgen naturalmente en el proceso de toma de decisiones cuando un conjunto de recursos ilimitados debe ser utilizado de la manera más efectiva para diferentes entidades. Entre las principales aplicaciones de estos modelos se encuentran la planeación

logística y de producción, planeación financiera, planeación de campañas de mercadotecnia y determinación de precios.

Los modelos de optimización matemáticos se utilizan cuando el proceso de toma de decisiones está en función de variables de decisión y el criterio que describe decisiones factibles puede ser expresado como un conjunto de igualdades o desigualdades matemáticas sobre las variables de decisión. Dentro de estos modelos están la optimización lineal, integral, convexa y de redes.

III.5.6 Modelos de manejo de proyectos: Están basados en la participación de varias disciplinas como la organización de negocios, psicología e investigación de operaciones. Los modelos matemáticos que entran dentro del manejo de proyectos juegan un papel muy importante en los procesos de toma de decisiones. Los modelos estocásticos o técnicas de evaluación y revisión de proyectos (*PERT*), son utilizados para derivar los tiempos de ejecución cuando se elaboran premisas estocásticas con respecto a la duración de las actividades, representadas por variables aleatorias.

III.5.7 Modelos de análisis de riesgo: Este tipo de análisis es particularmente útil cuando se necesitan tomar decisiones entre acciones que tendrán futuros complejos e inciertos en el futuro. El análisis de riesgo se aplica a una amplia gama de problemas de negocio como son la producción, el manejo de proyectos, inversiones, etc. principalmente basado en la teoría Bayesiana.

IV. LA INVESTIGACIÓN DE MERCADOS Y SUS APLICACIONES

IV.1 DEFINICIÓN E IMPORTANCIA DE LA INVESTIGACIÓN DE MERCADOS

La Investigación de Mercados es la función que une al consumidor, al cliente y al público en general, con el mercadólogo a través de información – información utilizada para identificar y definir oportunidades y problema en el mercadeo; generar, refinar y evaluar acciones de mercadotecnia; monitorear el desempeño de la mercadotecnia; y mejorar el entendimiento de la mercadotecnia como un proceso. ^[1]

Es también la identificación, recolección, análisis, diseminación y uso de manera sistemática y objetiva de información, con el propósito de ayudar a manejar la toma de decisiones relacionada con la identificación y la solución de problemas (y oportunidades) en el área de mercadotecnia. ^[2]

La toma de decisiones en el campo de la mercadotecnia implica aspectos que van desde cambios fundamentales en el posicionamiento de un negocio hasta la decisión de entrar a un nuevo mercado. El contexto en el que se toman estas decisiones es el proceso de planeación de mercados, conformado por cuatro etapas:

- **Análisis de la situación:** En esta etapa se entiende el entorno en el que se está trabajando así como el mercado en el que se participa; se identifican oportunidades y amenazas y se evalúa la posición de la competencia.
- **Desarrollo de estrategias:** Es en esta etapa en la que se define el alcance del negocio y los segmentos de mercado que serán atendidos; se establecen las ventajas competitivas y se fijan los objetivos internos de desempeño.
- **Desarrollo del programa de mercadotecnia:** Es aquí en donde se decide qué productos serán distribuidos y a través de qué canales (desarrollo de nuevos productos); se toman las decisiones de comunicación internas y externas con los clientes y consumidores (lanzamiento de una nueva campaña de publicidad); se fijan los precios y se toman las decisiones de venta personalizadas.
- **Implementación:** Se monitorea el desempeño del programa y se refinan las estrategias y el programa en general. Para la que la investigación sea eficaz en esta etapa, es importante que se establezcan objetivos específicos medibles para todos los elementos del programa de mercadotecnia.

La definición de investigación de mercados establece que las organizaciones participan en ella por dos razones: identificar y resolver problemas de mercadotecnia. Esta distinción sirve como base para clasificar la investigación de mercados en dos partes: investigación para identificar el problema e investigación para resolver el problema.

1. *Definición de la American Marketing Association*

2. *Marketing Research: An applied Orientation/ Malhotra, Naresh K./Prentice Hall, Inc./1996*

La figura IV.1 muestra esta clasificación y las tareas que la conforman:

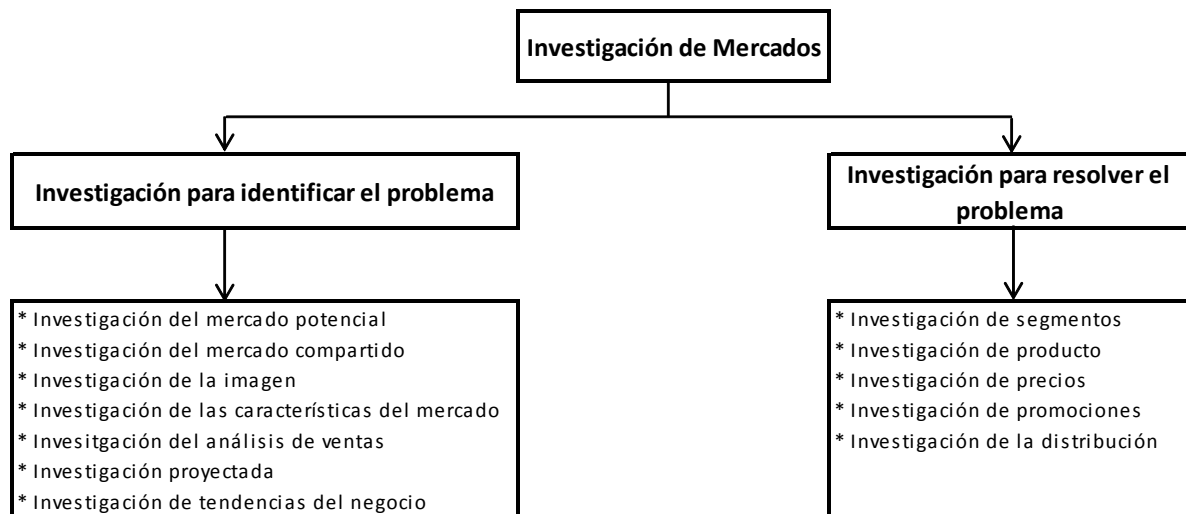


Fig. IV.1 Clasificación de la investigación de mercados

Sin embargo, ambas clasificaciones son necesarias para obtener resultados óptimos ya que una identificación correcta del problema derivará en una investigación óptima para la resolución del mismo. La investigación para resolver el problema es en donde las técnicas de minería de datos e inteligencia de negocios son aplicadas: en la parte de segmentación se tienen que determinar las bases para hacerlo, seleccionar mercados blanco con base en características demográficas, de imagen de producto, etc. La investigación de producto lleva a las pruebas del mismo con sus respectivas modificaciones, las pruebas en el mercado, el posicionamiento de las marcas y las pruebas de control en el almacenaje. Con respecto a la investigación de precios se establecen las políticas de los mismos, su comportamiento frente a la demanda y las técnicas de respuesta a los cambios en el mercado. En la investigación de las promociones se elaboran presupuestos para la creación de promociones óptimas, la relación con las ventas, se toman las decisiones de imagen, medios, temas creativos y se evalúa la efectividad de los anuncios. Finalmente la investigación de la distribución determina los canales de distribución y la localización de los puntos de venta y de los almacenes.

Los directivos y gerentes concuerdan en que la calidad técnica de la investigación es el principal determinante del empleo de la misma y toman en cuenta factores de conformidad con las expectativas previas, la claridad en la presentación de los resultados y la aceptación política dentro de la empresa. Tomando en cuenta los factores anteriores se podrá llevar a la empresa a la toma de mejores decisiones con bases sólidas y formales.

IV. 2 EL PROCESO DE INVESTIGACIÓN DE MERCADOS

La investigación de mercados es vista entonces como un proceso compuesto por una serie de pasos, los cuales definen las tareas que deben ser cumplidas al llevar a cabo un estudio de investigación de mercados. Estos pasos incluyen la definición del problema, el desarrollo de una aproximación a dicho problema, la formulación del diseño de investigación, trabajo de campo, preparación y análisis de los datos y finalmente, la generación y presentación de reportes.

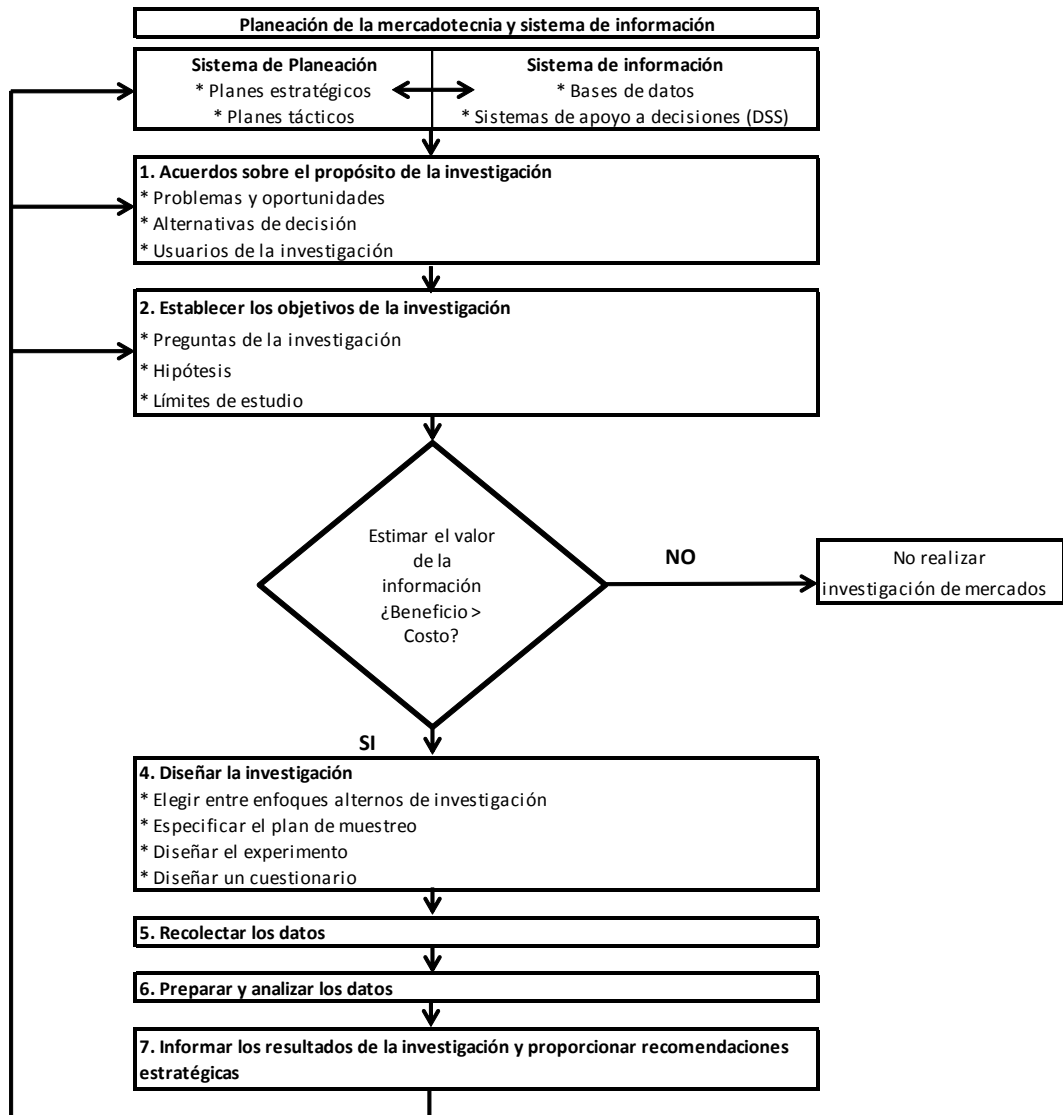


Fig. IV.2 El proceso de investigación de mercados

La figura IV. 2 muestra de manera detallada el proceso de investigación de mercados y en qué parte del mismo, las bases y almacenes de datos así como los sistemas de apoyo a las decisiones, juegan un papel determinante.

En el primer paso **Acordar sobre el propósito de la investigación**, el directivo o gerente y el encargado de realizar la investigación, deben tener la misma opinión sobre los problemas y oportunidades en el futuro inmediato, las alternativas de decisión que se deben evaluar y si la toma de decisiones será realizada de manera unilateral o dentro de un grupo.

En el segundo paso **Establecer los objetivos de la investigación**, se define qué información se necesita. Es en este paso en donde el o los encargados de la toma de decisiones deben especificar el tipo de conocimiento que desean adquirir para después desarrollar una hipótesis a demostrar y finalmente definir el alcance que se tendrá al realizar la investigación.

Una vez obtenida la información llega el momento de tomar la decisión y analizar si el valor de la información llevará a que el beneficio que se obtendrá será mayor al costo. Si no es así, es preferible no continuar con el proceso de investigación de mercados y redefinir el propósito de la investigación o los objetivos de la misma.

Cuando se ha llegado a la conclusión de que la investigación de mercados tendrá un beneficio mayor al costo de realizarla, se procede a **diseñar** las estrategias para hacerlo. En este punto se elige el método para recopilación de datos que puede ir desde métodos cualitativos hasta encuestas y experimentos; se procede al desarrollo de los mismos así como a la elaboración de un plan de muestreo y análisis.

Ya con el diseño de la investigación listo llega el momento de **Recolectar los datos**, mediante trabajo de campo, encuestas telefónicas y recepción de información por parte de los clientes y consumidores. Estos datos tendrán que ser almacenados en bases de datos para la utilización posterior de los mismos.

Cuando los **datos** han sido recopilados es necesario **prepararlos** para eliminar la información que pueda llevar a resultados equivocados y entonces se procede al **análisis** de los mismos utilizando las técnicas que ya se han descrito, como las pruebas de hipótesis, histogramas, medidas estadísticas (media, mediana, moda), análisis de correlación y de regresión, análisis de grupos o *clusters*, etc.

Finalmente es necesaria la **presentación de la información** obtenida mediante la elaboración de reportes ejecutivos u operativos que puedan permitir la toma de nuevas decisiones o la adecuación en procesos o estrategias dentro de la organización.

Si la información no satisfizo las necesidades de las áreas que la solicitaron, entonces se deben revisar el propósito o los objetivos de la misma, así como realizar nuevas solicitudes de información dentro del almacén de datos de la empresa que ayuden a obtener resultados más valiosos para tomar decisiones de manera inteligente.

IV. 3 EL DATABASE MARKETING

Como se vio en el capítulo II, el que la empresa cuente con un almacén de datos en el que se vaya registrando el comportamiento histórico de los clientes ayudan a la creación de estrategia y análisis que aumentan su posibilidad de éxito frente a sus competidores. La suma del manejo de estas bases de datos con el empleo de modelos de mercadotecnia da como resultado las prácticas denominadas *Database Marketing* o mercadotecnia de bases de datos

El *Database Marketing* puede:^[3]

- Lograr empatar los productos o servicios con los deseos y las necesidades de los clientes
- Ayudar a seleccionar nuevas listas o a usar nuevos medios de publicidad que se ajusten al perfil de los clientes actuales
- Maximizar la personalización de todas las ofertas para cada cliente.
- Proporcionar una interacción continua con los clientes y los prospectos.
- Señalar el momento y la frecuencia ideales para las promociones.
- Medir la respuesta y explicar los resultados.
- Ayudar a crear las ofertas que probablemente generen más respuestas de los clientes.
- Ayudar a lograr una propuesta de venta única, dirigida a atraer a nuevos clientes.
- Integrar la comunicación de respuesta directa con otras formas de publicidad.
- Demostrar que los clientes son activos valiosos para la empresa.

Con una base de datos, los mercadólogos pueden usar las acciones anteriores de los clientes para predecir sus preferencias futuras o para crear el perfil de clientes o prospectos para una segmentación de mercados eficaz. Asimismo, pueden proyectar ventas adicionales y apuntar eficazmente hacia nuevos clientes con las mismas características, e incluso predecir el valor de por vida de los clientes recientemente adquiridos.

Para que una base de datos sea manipulable y al mismo tiempo ayude a la obtención del conocimiento necesario para la toma de decisiones, al menos debe tener los siguientes campos:

- Identificador único: número de cliente, número de socio, número de contrato.
- Nombre completo del individuo: apellidos, nombre, título.
- Dirección de correo: calle, número, C.P., ciudad, estado, país
- Número de teléfono: fijo, celular, oficina
- Fecha y detalles de las primeras transacciones: fecha de venta, fecha de alta en el sistema, fecha de primer pago, fecha primera de utilización.
- Historial de transacciones: última transacción, frecuencia, valor monetario por fecha, importes por transacción, importes acumulados.
- Historial crediticio o algún tipo de clasificación: cliente frecuente, cliente VIP.
- Datos demográficos: edad, sexo, estado civil, datos de los familiares, ingresos, ocupación.
- Fechas de vencimiento: de pago, de contrato, de membresía

3. *How to find and cultivate customers through direct marketing/ Baier, Martin/Lincolnwood NTC*

Existen muchos otros datos que se pueden almacenar dentro de la base y bien, su número, podría tender a ser infinito. Es aquí donde los encargados de la toma de decisiones deben valorar si cierto dato solicitado será útil en un futuro o si tendrá un valor agregado para la organización y los análisis que se realicen. Muchas veces, las áreas operativas que están en contacto con los clientes, piensan que entre más datos haya en la base, sus actividades serán más exitosas. Lo que las áreas administrativas, de análisis y de sistemas tienen que puntualizar y demostrar es que no todos los datos sirven en un ambiente de inteligencia de negocios; lo que se necesita es realizar análisis profundos que arrojen la información precisa para mejorar el desempeño de todas las áreas de la organización.

Algunos otros datos que dan un valor agregado a las bases son:

- Histórico de transacciones
- Cambios de datos personales: dirección, teléfono, ciudades, países.
- Idioma de contacto

Otra opción es definir varios tipos de bases de datos dependiendo de las transacciones que cada una contenga y después realizar los cruces necesarios entre ellas para encontrar patrones de comportamiento que arrojen datos valiosos sobre los clientes.

La recomendación es que las empresas tengan bases de clientes y prospectos por separado, para evitar cruces de información o dificultades en el manejo de los datos. Es necesario también clasificar a los clientes con el fin de identificar los clientes “tipo” y así ser más eficaces en la creación de prospectos; identificar mejor a los mejores clientes; identificar mercados nicho para agregarlos al universo de mercadotecnia; y desarrollar herramientas de mercadotecnia más eficaces. Con base en esto, lo ideal es dividir a los clientes en activos, inactivos y consultas. Los clientes activos son los que han comprado o utilizado el producto recientemente, es preciso conocer la frecuencia de la utilización, el monto invertido y sus preferencias; de esta manera se pueden designar los recursos a este segmento de clientes que es más rentable. En cuanto a los clientes inactivos, es necesario identificar el tiempo que llevan sin consumir, cuál fue su utilización mientras estuvieron activos, qué tipo de acercamiento se ha tenido con ellos desde entonces; con base en esto se pueden diseñar promociones específicas para este tipo de clientes. Las consultas son aquellas personas que no son ni han sido clientes pero que se han acercado para conocer el producto, y lo ideal es obtener datos que puedan convertir estas consultas en prospectos y por ende, en clientes potenciales.

Es un hecho que contactar a cada uno de los miembros de una base de datos puede ser muy costoso. El punto de equilibrio que se necesita en una tasa de respuesta para que el contacto sea rentable es:^[4]

$$\text{Punto de equilibrio de la Tasa de respuesta} = \frac{\text{Costo de contacto}}{\text{Ingresos netos esperados de un entrevistado}}$$

4. *Marketing Research/Aaker, Kumar, Day/Limusa Wiley/2001*

Así, una regla importante es abordar sólo a aquellos segmentos cuyas tasas de respuesta esperadas sean superiores a la tasa del punto de equilibrio

Otra herramienta esencial para identificar a los mejores clientes es la fórmula R/F/M. Donde:

- R = Novedad. El tiempo que ha pasado desde que el cliente utilizó el producto o realizó la compra.
- F = Frecuencia: Frecuencia de la utilización
- M = Valor monetario de la transacción.

Aquí interviene la decisión de los encargados de tomar las decisiones sobre cuál es la importancia relativa que se da a cada una de las variables.^[5] Así para obtener este tipo de calificación de los clientes, cada empresa puede asignar puntos específicos para cada rango de novedad, frecuencia y valor monetario, para después ponderarlos entre todos los clientes y finalmente obtener los puntos acumulados que nos ayudarán a identificar quiénes son los mejores.

El *Database Marketing* ayuda a la retención de los clientes bajo la premisa de que los clientes son más fáciles de retener que de adquirir. Si todos los esfuerzos de la organización están dirigidos a la obtención de nuevos clientes y no se preocupan por la retención de los ya existentes, se tendrán pérdidas monetarias y esfuerzos inútiles. La regla de Pareto descrita en el capítulo III.4 como método de control, se puede aplicar también en la distribución de los esfuerzos ya que el 80 por ciento de los ingresos del negocio proviene del 20 por ciento de los clientes.

Es útil también determinar el valor de los clientes durante su ciclo de vida, es decir, cuál es el costo-beneficio neto de un cliente al mantenerlo dentro de la cartera por un determinado periodo de tiempo.

Sin duda, tomando en cuenta estos puntos, los resultados serán superiores, se tendrán clientes y colaboradores satisfechos y sobre todo, se podrán disminuir los costos y aumentar los ingresos de la empresa.

5. *The Complete Database Marketer/ Hughes, Arthur/Chicago: Irwin*

IV.4 LA MERCADOTECNIA DE RELACIONES

El desarrollo de relaciones con los clientes, comprendiendo sus gustos y preferencias en forma individual, da origen a la mercadotecnia de relaciones que combina elementos de publicidad general, promoción de ventas, relaciones públicas y mercadotecnia directa para crear formas más eficaces y eficientes de llegar a los clientes. Basándose en la mercadotecnia de relaciones, las campañas que se lancen serán cada vez más específicas y atacarán necesidades cada vez más puntuales de cada uno de los clientes con lo que también, se podrá dividir a los encargados del contacto con ellos con base en tareas y objetivos bien establecidos.

La habilidad de explotar efectivamente la información reunida sobre el comportamiento de los clientes, representa hoy en día una poderosa arma competitiva para las empresas. Una compañía capaz de reunir, almacenar, analizar y entender la gran cantidad de datos sobre sus clientes (Inteligencia de Negocios) puede basar sus acciones de mercadotecnia en el conocimiento extraído y alcanzar ventajas competitivas sustentables. Las empresas pueden adoptar estrategias basadas en la mercadotecnia de relaciones para transformar contactos ocasionales con los clientes en relaciones personalizadas a largo plazo. De esta manera, es posible alcanzar una mayor satisfacción de los clientes y, al mismo tiempo, elevar las ganancias de la compañía, en un ambiente de beneficios mutuos.

La figura IV.3 muestra claramente la red de relaciones que se debe establecer dentro de la organización, centrando los esfuerzos de cada una de las áreas y tareas, en las estrategias de mercadotecnia de relaciones.

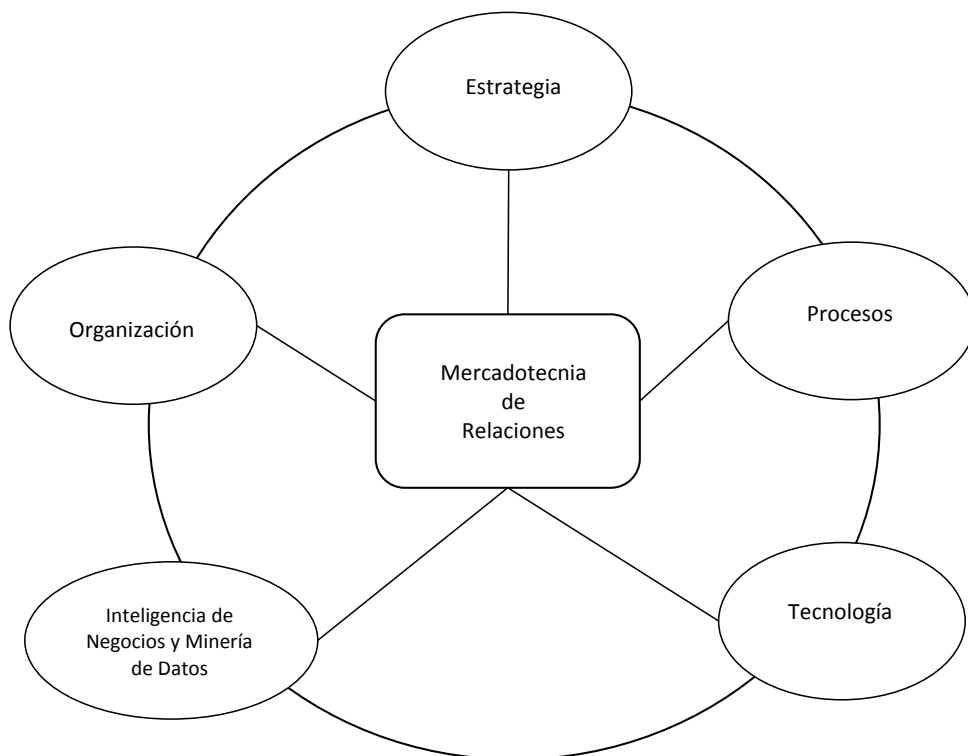


Fig. IV.3 Componentes de una estrategia de mercadotecnia de relaciones

Para obtener las ventajas deseadas, una compañía debe adoptar las estrategias de mercadotecnia de relaciones como un proyecto donde los distintos departamentos de la misma están encaminados a integrar una cultura organizacional y de recursos humanos que cooperen entre sí, para ello, es necesario crear, dentro de la compañía, una verdadera cultura de los datos, como ya se ha visto en los capítulos I y II.

Los empleados, proveedores, fuerza de ventas y colaboradores en general, deben dirigir sus esfuerzos a desarrollar una visión empresarial que ponga a los clientes en el centro de toda la organización como se muestra en la Figura IV.4, con ello se logra un intercambio de mutuo beneficio entre todos los sujetos involucrados.

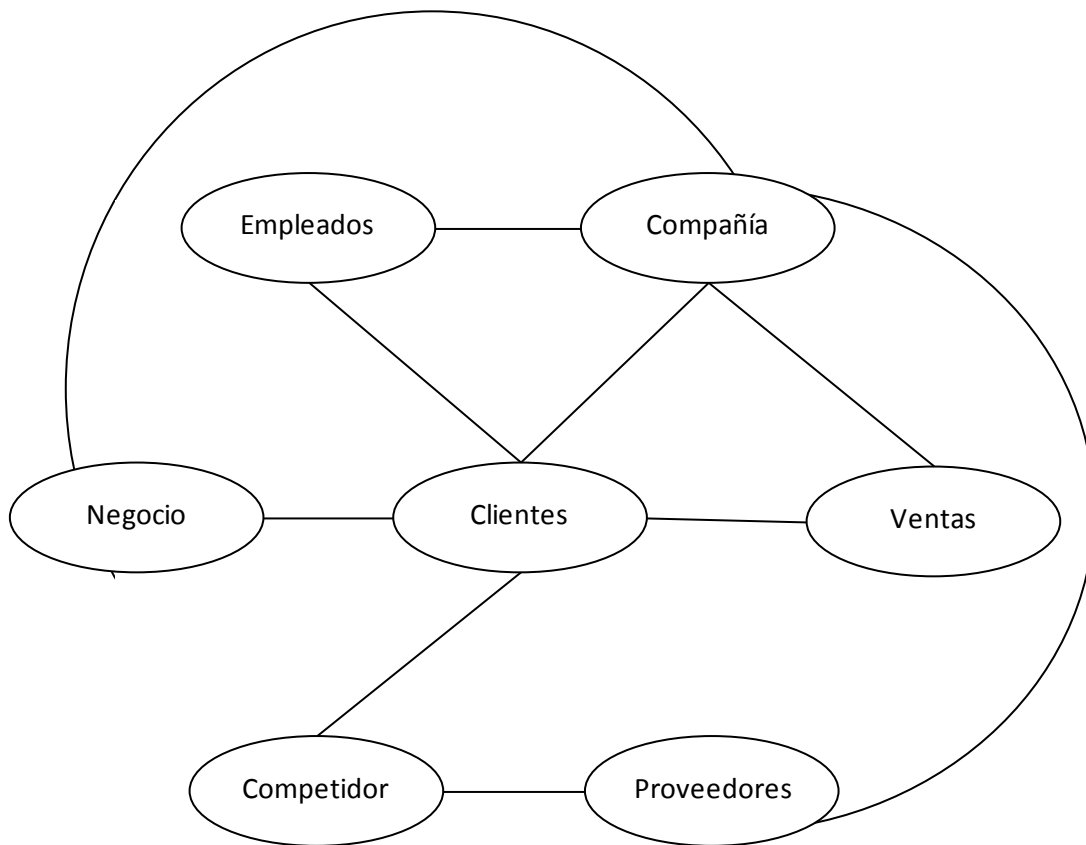


Fig. IV.4 Red de relaciones establecidas en una estrategia de mercadotecnia de relaciones

Existen varias reglas con respecto a la mercadotecnia de relaciones que se pueden analizar y establecer y que sirven para mejorar la comunicación con los clientes, economizar costos, mejorar la eficiencia en ventas y personalizar los productos. Por ejemplo:

- Las relaciones empresa-cliente:

- Las relaciones de mayor intensidad se establecen con pocos clientes que representan un alto valor para la empresa. Estas relaciones son las llamadas *business-to-business* (B2B) o “negocio para negocio” cuyo nombre se refiere a que, las organizaciones utilizan recursos dedicados especialmente en mantener y fortalecer este tipo de relaciones, es decir, utilizan los elementos del negocio para incrementar el negocio.
- Las relaciones de menor intensidad se establecen con la mayoría de los clientes que representan un valor bajo para la empresa. Estas son las relaciones *business-to-customer* (B2C) o “negocio dirigido al cliente” y engloba la gran cantidad de contactos que se establecen con clientes que no tienen un valor muy importante para el negocio y que se comunican con la empresa de una manera impersonal, a través de ejecutivos de atención telefónica, páginas web y puntos de venta. Los análisis de minería de datos para la segmentación y elaboración de perfiles son muy valiosos en estos casos, que se caracterizan por un gran número de contactos y transacciones.
- La relación costo de ventas y sus correspondientes ingresos:
 - Cuando las transacciones tienen una ganancia menor por unidad vendida, es necesario implementar acciones de bajo costo, como en los esfuerzos de mercadotecnia masiva: correo directo, email marketing, etc.
 - Conforme la ganancia aumenta, se deben implementar estrategias más desarrolladas e intensas como se hace con una red de ventas directas.
- Finalmente, otra relación es la que se crea entre el grado de personalización que puede tener un esfuerzo de mercadotecnia y la complejidad de la entrega de una pieza y sus costos:
 - Si se quiere una entrega rápida, implicará disminución en los costos pero no un grado de personalización muy elevado, es decir, se elaborarán piezas masivamente.
 - Conforme crece el grado de personalización, el tiempo de entrega será más lento y se incurrirá en mayores costos.

Los análisis de minería de datos son una herramienta muy poderosa utilizada en la mercadotecnia de relaciones que ayudan a identificar los segmentos a los cuales deben enfocarse los productos que han sido personalizados. Es por esto que la mercadotecnia de relaciones, está íntimamente ligada a la inteligencia de negocios y a la minería de datos que forman parte de lo que se conoce como ambiente para los análisis de mercadotecnia de relaciones.

La figura IV.5 muestra los componentes de este ambiente y cuáles de estos forman parte de los procesos de toma de decisiones y cuáles de los sistemas de información:

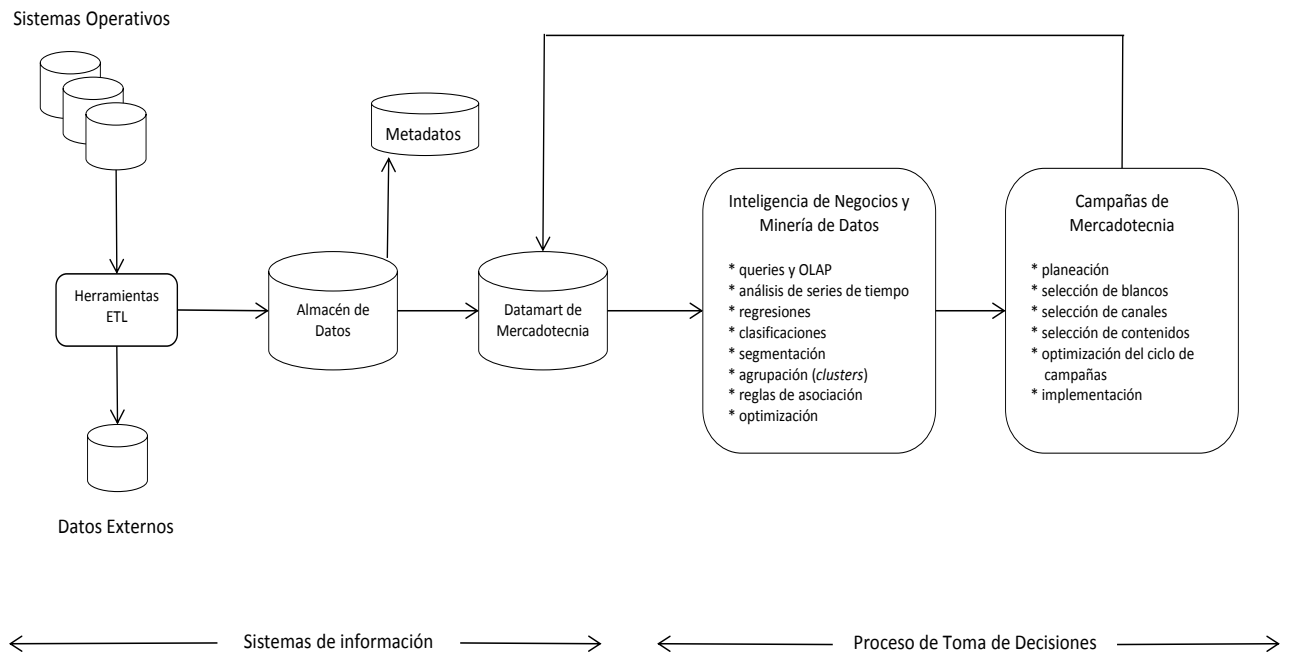


Fig. IV.5 Componentes que intervienen en un ambiente para el análisis de mercadotecnia de relaciones

Con esta figura se puede ver claramente cómo recopilando los datos externos y los datos incluidos en los sistemas operativos de la organización, se conforma el almacén de datos maestro, en el cual se tiene la información completa de los clientes, productos, servicios, pagos, uso, etc. De este almacén de datos principal se puede extraer el subconjunto de datos que ayudarán a las áreas de mercadotecnia, manipulando la información con base en la inteligencia de negocios y utilizando las herramientas de minería de datos: consultas, análisis de series de tiempo, regresiones y demás métodos que se vieron en el capítulo II.

Con base en los resultados obtenidos en estos procesos, las áreas de mercadotecnia pueden entonces planear sus campañas, elegir los blancos que serán más sensibles a ellas, las vías o canales por las cuáles se comunicará a los clientes de la existencia de estas campañas, etc. Finalmente los resultados de las campañas (redención, tipo de clientes que respondieron, utilización, etc.) serán nuevamente depositados en el *datamart* de mercadotecnia para futuras campañas y análisis, es decir que, el conocimiento de los perfiles de los clientes se utiliza para diseñar acciones de mercadotecnia específicas que son convertidas después en campañas promocionales.

Como recomendación, siempre es conveniente elegir, durante la ejecución de cualquier campaña, un grupo de control de clientes a los cuales no se dirigirá la acción específica de la misma; esto ayudará en la medición de la efectividad.

IV.5 INTELIGENCIA DE MERCADOTECNIA

Los procesos de decisión de mercadotecnia están caracterizados por un alto nivel de complejidad debido a la presencia simultánea de objetivos múltiples y las innumerables alternativas de acción, resultantes de la combinación de la gran cantidad de opciones a elegir disponibles para los encargados de la toma de decisiones.

Como consecuencia de esta gran cantidad de opciones se han desarrollado numerosos modelos matemáticos aplicados a la mercadotecnia y al proceso de ventas, lo que se conoce hoy en día como Inteligencia de Mercadotecnia, que consiste en la aplicación de modelos predictivos para apoyar las estrategias de mercadotecnia de relaciones, cuyo propósito, como ya se ha visto es el de personalizar y fortalecer la relación entre una empresa y sus clientes.

Para saber qué tipo de modelos y análisis se deben realizar, es importante conocer el ciclo de vida de un cliente cuyo valor se va modificando durante su etapa como tal en la organización. Estas etapas se pueden dividir como sigue:

- **Prospectos:** son los clientes potenciales o las personas que no han adquirido aún el producto. En este punto se deben llevar a cabo acciones de adquisición directas (llamadas telefónicas, emails, venta directa) e indirectas (anuncios, información vía páginas web). Estas acciones conllevan un costo que puede ser asignado a cada prospecto y conformar una pérdida económica acumulada para la empresa que durará hasta que dicho prospecto se convierta en cliente.
- **Nuevos clientes y clientes regulares:** a partir de que un prospecto decide adquirir el producto, se tienen que considerar acciones que garanticen la estancia de dicho cliente dentro de la organización lo que llevará a que el nivel de pérdida económica disminuya hasta convertirse en ganancias. Para estos nuevos clientes se deben de llevar a cabo actividades de venta cruzada (ofreciendo productos adicionales a los adquiridos), mejora de venta (ofreciendo productos internos de un mejor nivel y que representan una mayor ganancia para la empresa) y retención.
- **Desertores:** son los clientes que dan por terminada de manera voluntaria (porque no le satisfizo el servicio o los productos y busca otras opciones), forzada (porque, por ejemplo, el cliente no ha cumplido con los pagos) o no intencional (porque el cliente se muda de domicilio o fallece)

Las estrategias de mercadotecnia de relaciones pueden mejorar la lealtad de los clientes, ayudando a extender su relación con la empresa y por lo tanto, mejorando las ganancias de las mismas.

Las herramientas de minería de datos ayudan a elaborar perfiles de adquisición y retención de los clientes, a que las ventas estén enfocadas a los clientes de manera personalizada, a realizar compras en línea, a predecir el tamaño del equipo requerido para realizar las ventas o conformar un centro de atención telefónica, a seleccionar un mercado objetivo para determinados productos y para realizar labores de minería web para buscar nuevos prospectos.

El análisis de los clientes con una mayor lealtad comparado con los clientes que desertaron de la empresa de manera voluntaria puede ayudar a crear campañas de retención dirigidas a los segmentos importantes que representen clientes de alto valor y que tienen un alto riesgo de desertación. Para obtener un sistema de calificación que indique la probabilidad de desertación de cada cliente, es necesario realizar una segmentación basada en los datos históricos de esta variable. El hecho de predecir el riesgo de desertación requiere el análisis de los registros de las transacciones de cada cliente y la identificación de los atributos más relevantes que expliquen con mayor precisión este hecho.

Otro análisis es el de los clientes que son potencialmente sensibles a adquirir mejores productos o productos adicionales para también dirigirles esfuerzos específicos. Mediante modelos de clasificación, es posible identificar a los clientes que tienen una mayor probabilidad a adquirir una oferta de venta cruzada o una mejora en los que actualmente tienen.

El análisis de enlaces o conexiones o *Market Basket* que se mencionó en el capítulo II, es otro de los análisis que apoyan a la inteligencia de mercadotecnia, sobre todo en la industria de ventas al menudeo y en el comercio electrónico ya que no se tienen conocimientos profundos de los clientes que están adquiriendo un determinado producto. El propósito de este análisis es extraer información poderosa que ayude a planear estrategias de mercadotecnia. Los datos utilizados en este caso se refieren a las transacciones de compra y pueden ser asociados con el tiempo si el comprador puede ser identificado con una tarjeta de crédito o una factura. Cada transacción consiste en una lista de objetos comprados. Esta lista es llamada canasta o *basket*. Con base en esta lista se crean correlaciones que aportan información al proceso de toma de decisiones y por ellos se pueden ofrecer indirectamente productos adicionales a los que se están comprando pero que son, compras potenciales por parte de los clientes.

Una vez más aparecen los modelos de agrupación o *clusters* como una importante herramienta de inteligencia de mercadotecnia y que identifican grupos homogéneos a los cuales se dirigirán esfuerzos específicos que, se sabe, por las similitudes del comportamiento de los clientes, tendrán una buena redención.

A continuación se presentarán algunos modelos para diseñar y planear un plan de mercadotecnia y ventas como son las funciones de respuesta, y la planeación de llamadas y de recursos para las ventas.

- a) **Funciones de respuesta:** Describen la elasticidad de las ventas en términos de la intensidad de las acciones de venta y son un método formal para describir las complejas relaciones existentes entre dichas acciones y las reacciones del mercado. Las funciones de respuesta están expresadas en términos de ingresos por ventas. La intensidad de una acción de venta puede relacionarse con diferentes variables como son el número de llamadas a un cliente en un periodo determinado, el número de menciones que se hacen de un producto en cada periodo y el tiempo dedicado a un cliente en cada periodo. Las funciones de respuesta pueden ser estimadas considerando dos tipos de información. Primero, los datos disponibles del pasado con respecto a la intensidad de las actividades que fueron llevadas a cabo y sus correspondientes ventas para desarrollar un modelo de regresión paramétrico a través

de variantes de modelos de regresión. Segundo, la experiencia directa de los agentes telefónicos y de ventas para obtener información subjetiva que pueda ser después incorporada a la función de respuesta.

Sea entonces x_h el número de llamadas que hace un agente específico a un cliente en un periodo h de tiempo y $r_h(x_h)$ el valor de ventas del cliente h asociado con el x_h . La variable que determina la función de respuesta r expresa la intensidad de la actividad de ventas que se llevó a cabo. Una forma paramétrica debe ser seleccionada para expresar esta dependencia funcional, lo que resulta en:

$$s_h(x_h) = r_0 + (r_\infty - r_0) \frac{x_h^\sigma}{\gamma + x_h^\sigma}$$

Donde r_0 representa el nivel de ventas que se hubiera obtenido con una actividad de ventas igual a 0, como un efecto prolongado de acciones previas; r_∞ representa el nivel máximo de ventas con independencia de la intensidad de la actividad de ventas; γ y σ son dos parámetros que deben ser estimados.

Para obtener una estimación de los cuatro parámetros que aparecen en la expresión $r_h(x_h)$ es posible proceder de dos maneras complementarias. Los datos de ventas históricos, se pueden utilizar para crear un modelo de regresión y determinar los valores a través del método de mínimos cuadrados. Para incrementar el valor de las opiniones de los agentes de ventas, también es posible preguntar a los agentes y gerentes de ventas para estimar los parámetros r_0 y r_∞ , así como también los valores de las ventas esperadas en otros tres puntos críticos de las funciones de respuesta: $r(x_h)$, que corresponde al número de llamadas que son llevadas a cabo en el momento del análisis, $r(\frac{1}{2}x_h)$ y $r(\frac{3}{2}x_h)$, asociados respectivamente con el incremento y decremento del número de llamadas en un 50% con respecto al valor actual. Basados en una evaluación subjetiva de los cinco valores de respuesta derivados de este proceso, se puede obtener una estimación por interpolación de la escala de parámetros γ y σ .

- b) Planeación de llamadas:** El objetivo de estos modelos de optimización se es proporcionar a cada agente telefónico, un plan de actividades óptimo, para esto se debe identificar el número óptimo de llamadas para cada cliente o prospecto de una asignación específica de cartera para un agente. La función objetivo, expresa la diferencia entre los ingresos y los costos de transferencia.

Las variables de decisión se definen como sigue:

$X_h =$ Número de llamadas para un cliente o prospecto h ,

$W_j =$ Número de viajes al área de mercado j ,

Mientras que los parámetros tienen los siguientes significados:

- $a_h =$ Relevancia estratégica de un cliente o prospecto h ,
 $c_j =$ Costo de transferencia al área j ,
 $v_j =$ Tiempo de transferencia al área j ,
 $t_h =$ Tiempo empleado en un cliente o prospecto h en cada llamada,
 $l_h =$ Número mínimo de llamadas a un cliente o prospecto h ,
 $u_h =$ Número máximo de llamadas a un cliente o prospecto h ,
 $b =$ Tiempo total disponible para el agente telefónico.

El correspondiente problema de optimización se puede formular entonces como sigue:

$$\max \sum_{h \in H} a_h r(x_h) - \sum_{j \in J} c_j w_j$$

s.to
$$\sum_{h \in H} t_h x_h + \sum_{j \in J} v_j w_j \leq b$$

$$X_h \leq u_h, X_h \geq l_h, \quad h \in H, \text{ donde } H \text{ es el número total de prospectos o clientes}$$

$$W_j \geq X_h, \quad j, h \in D, \text{ donde } D \text{ es el conjunto de clientes localizados en un área específica.}$$

$$X_h, W_j \geq 0, \text{ e integral,} \quad h \in H, j \in J.$$

La segunda regla expresa un límite en el tiempo total disponible para el agente telefónico dentro del horizonte de planeación. La tercera regla impone un límite superior y uno inferior, respectivamente en el número de llamadas para cada cliente o prospecto. Finalmente la última regla, establece una condición de consistencia lógica entre las variables de decisión X_h y W_j .

Para obtener una solución al modelo se puede proceder como sigue: primero la función de respuesta se aproxima con una función lineal piecewise, derivando un conjunto de problemas de optimización de integrales lineales mixtas. Estas se resuelven utilizando el método de *branch-and-bound*, posiblemente truncado para limitar el tiempo de cálculo y alcanzar soluciones cuasi óptimas. Una alternativa de solución, es un algoritmo de aproximación.

c) **Modelo general para la planeación de los recursos de venta:** Para este caso se puede adoptar un marco de representación que involucre el elaborar una lista de todas las tareas que pueden ser asignadas a un agente. El modelo resultante, da como resultado el plan óptimo para los agentes a través de múltiples periodos de tiempo, tomando en cuenta distintos recursos compartidos.

Para cada agente i y para cada periodo t , el conjunto de todas las diferentes actividades de venta, que representan el plan de llamadas y presentación de productos a diferentes clientes, se identifica como S_{it} . Los recursos requeridos, se denotan por $g \in \mathcal{G} = \{1, 2, \dots, G\}$, representan el presupuesto general disponible para implementar dichas acciones de venta, u otros factores técnicos necesarios para adoptar las diferentes acciones.

Las variables binarias de decisión requeridas Y_{iut} se definen como:

$$Y_{iut} = \begin{cases} 1 & \text{si la acción } u \in S_{it} \text{ se selecciona para el agente } i \text{ en el periodo } t, \\ 0 & \text{si no se selecciona.} \end{cases}$$

Los parámetros son:

S_{it} = Conjunto de acciones factibles para el agente i en el periodo t ,

w_{giut} = Cantidad del recurso g requerida para implementar la acción $u \in S_{it}$ por el agente i en el periodo t ,

V_{gt} = Cantidad del recurso g disponible en el periodo t ,

v_{iut} = Valor de ganancia asociado con la acción $u \in S_{it}$.

El problema de optimización resultante es:

$$\max \sum_{t \in T} \sum_{i \in I} \sum_{u \in S_{it}} v_{iut} Y_{iut}$$

$$\text{s.to} \quad \sum_{i \in I} \sum_{u \in S_{it}} w_{giut} Y_{iut} \leq V_{gt}, \quad g \in G, t \in T,$$

$$\sum_{u \in S_{it}} Y_{iut} = 1, \quad i \in I, t \in T,$$

$$Y_{iut} \in \{0,1\}, \quad i \in I, u \in S_{it}, t \in T$$

La primera regla expresa el límite superior de la cantidad disponible de cada recurso en cada periodo. La segunda regla representa una condición de elección múltiple impuesta para garantizar que cada agente va a llevar a cabo exactamente una acción en cada periodo.

El modelo finalmente es un problema de optimización binario que pertenece a la clase de problemas de *knapsack* de opción múltiple generalizados. Se recomienda reducir con anticipación el número de acciones disponibles para cada agente, por medio de una fase de pre procesamiento encaminada a descartar aquellas acciones que se observan cómo menos convenientes y que generarán menores ganancias.

V. CASO PRÁCTICO: IMPLEMENTACIÓN DE LA INTELIGENCIA DE NEGOCIOS EN RHC

V.1 ANTECEDENTES HISTÓRICOS Y OPERATIVOS

Royal Holiday Club (RHC) es una empresa mexicana fundada en el año de 1983 y que introduce al mercado una alternativa al concepto vacacional clásico de Tiempo Compartido. En el año de 1988, lanza una nueva opción de membresía vacacional llamada “*multi-resort*” en donde, los clientes del negocio, no tienen la obligación de pasar sus vacaciones siempre en el mismo destino y hotel, sino que se les ofrecen varias opciones de acuerdo a sus gustos y necesidades. En 1990, como parte de la expansión de la empresa, se enriquece el concepto de “*destinos flota*” ofreciendo más de veinte opciones diferentes para viajar.

Anticipándose a las necesidades del mercado turístico, en 1993, se crea un innovador y único concepto basado en un sistema de **puntos**, llamados **Créditos Vacacionales** (*Holiday Credits*). A partir de este momento, los socios del club vacacional comienzan a tener mayores opciones para sus vacaciones, ya que pueden intercambiar los créditos de la membresía que adquirieron por noches de alojamiento en hoteles propios y hoteles asociados, por días en cruceros y por servicios de viaje adicionales como son boletos de avión, seguros de viajero, atracciones, etc. Con este nuevo sistema se logra expandir el negocio a destinos como América del Sur, África y Medio Oriente en el año de 1995.

En el año 2000, Royal Holiday es el primer Club Vacacional que ofrece su propio programa de cruceros con todas las navieras alrededor del mundo. Después de 20 años en el mercado, a partir del año 2003, la empresa comienza a adquirir créditos con bancos e instituciones financieras en México, Estados Unidos y Brasil, con lo que la deuda interna de la empresa se vuelve un tema importante para la operación y el manejo de flujo de efectivo, que sólo podría ser cubierta a través de los ingresos que las reservaciones y cobranza a los socios produjeran; esto se vuelve crítico después de los huracanes del año 2006 que dejan serios daños en los hoteles propios y en los que se tienen que invertir fuertes sumas de capital para su recuperación. Las ventas, que hasta 2008 tuvieron una tendencia creciente, servirían entonces para la apertura de nuevos puntos de venta en América del Sur, Puerto Rico, Bahamas y República Dominicana.

La crisis mundial de los años 2008 y 2009, tuvo grandes repercusiones en los temas de uso de la membresía y morosidad de pago por parte de los socios, por lo que fue necesario elaborar planes de contingencia para incentivar ambas operaciones. Los resultados no fueron tan exitosos como se hubiera esperado puesto que fueron acciones reactivas y no preventivas, debido a que la estructura de la empresa no está basada en un ambiente de Inteligencia de Negocios. Para resolver este tema, en el año 2009 se propone la implementación del sistema SAP en toda la empresa: oficinas corporativas en la Ciudad de México, salas de ventas y oficinas foráneas; así como la creación de una plataforma hecha a la medida de la empresa para satisfacer las necesidades operativas de cobranza, servicio (reservaciones) y ventas.

Hasta el año 2012 no se ha logrado lanzar oficialmente dicho sistema ni tampoco el ERP con SAP dado que las decisiones que se han tomado en la empresa no han estado basadas, en su mayoría, en un análisis profundo de las bases de datos, de los reportes y del enfoque a

mejoras en el desempeño de la empresa en general, sino a las necesidades básicas, del día a día de las áreas operativas.

La empresa cuenta con un sistema operativo global llamado **RCC**, diseñado especialmente para sus necesidades y en donde se ingresa toda la información de los prospectos y nuevas ventas, de los socios regulares (datos personales, uso, histórico de pagos) y que además, guarda información de todas y cada uno de los contactos que se tienen con los socios. Toda esta información se encuentra dividida en módulos y cuenta con historia desde 1983.

A pesar de contar con un importante equipo de sistemas (IT) que trabaja con dicho almacén de datos tan robusto y con sus respectivos *data marts* históricos de uso, reservaciones, mercadotecnia, ventas, etc., el análisis que se hace sobre ellos, aún tiene muchas deficiencias porque se hace de modo tradicional y hasta obsoleto:

- El departamento solicitante que tiene que elaborar el análisis, pide al departamento de sistemas que explote la base de datos mediante consultas (*queries*) con ciertas premisas. El tiempo de respuesta normalmente depende de la carga de trabajo del área de sistemas, cuyo equipo de “*Database Marketing*” no se da abasto con todas las solicitudes, ya que tiene que atender otras funciones como soporte técnico en contingencias, mantenimiento a la página web de los socios e intraweb, mantenimiento a los medidores de desempeño en ventas, etc.
- Una vez que el área de sistemas ha enviado la información al área solicitante vía email o colocando los datos en un repositorio en los servidores, esta área (administración de cartera, tesorería, cobranza, marketing, inventarios, ventas) se dispone a “trabajar” con dichos datos que, no son enviados de otra manera, que mediante una hoja de cálculo en Excel. Muchas veces, dependiendo de la solicitud, se envía más de una hoja puesto que los registros cuando son históricos, rebasan el tamaño de una hoja de cálculo tradicional.
- El área solicitante, entonces trabaja con las herramientas que Excel proporciona: copiando y pegando, cortando, introduciendo fórmulas y, muchas veces incluso, depurando la información que puede venir con inconsistencias.
- Este trabajo manual, evidentemente, provoca que el margen de error sea elevado y que muchas veces haya que revisar y rehacer hasta más de dos veces el mismo análisis puesto que se han obtenido resultados distintos. Peor aún, cuando dos áreas distintas elaboran el mismo análisis y al compararlos, no hay coincidencia alguna.
- Si se logró obtener un resultado “aceptable” para los gerentes y directores de área, entonces se tienen que elaborar los reportes ejecutivos para mostrarse a la dirección general, financiera u operativa. Esto implica una segunda depuración de los datos que son “importantes” para que se puedan tomar decisiones futuras en cuanto a staff, presupuestos de gastos, inversiones, esfuerzos y metas de cobranza y servicio, etc.

Es un hecho que todas estas tareas no están agregando nada de valor ni hay tiempo para ningún análisis posterior realmente profundo que ayude a la correcta toma de decisiones sino que simplemente están sacando el trabajo del día a día bajo la presión del tiempo de respuesta que se está requiriendo.

A todo lo anterior, se tiene que sumar que el ERP con que actualmente se trabaja en la empresa, está totalmente desconectado del almacén de datos, es decir, el ERP se alimenta de manera manual con los reportes que elaboran cada una de las áreas operativas, lo que hace que el trabajo de las áreas administrativas y contables sea interminable. Sumado a esto, cuando se tienen que hacer presupuestos o evaluar el éxito financiero de una campaña de mercadotecnia enfocada a incrementar las reservaciones o de esfuerzos de cobranza especializados para aumentar los ingresos, los resultados y proyecciones no concuerdan con la realidad de la empresa, o con la “realidad” que presentan los reportes manuales que se elaboraron.

En resumen, RHC a pesar de contar con un almacén de datos y data marts específicos, no desarrolla sus actividades bajo una visión de Inteligencia de Negocios y por lo tanto, la Toma de Decisiones, no se hace de una manera óptima.

La figura V.1 muestra la arquitectura operativa de la empresa:

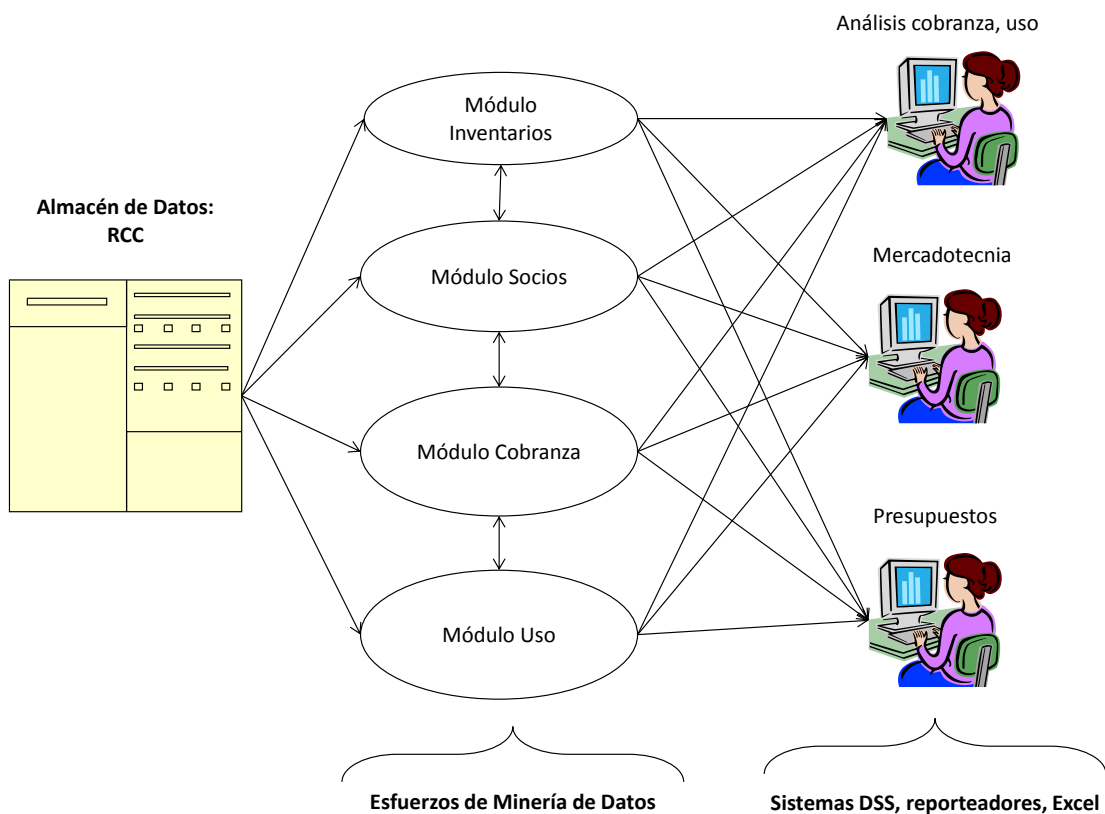


Fig. V.1 Arquitectura operativa en Royal Holiday Club

Se detecta entonces un gran proyecto de Inteligencia de Negocios debido a las siguientes premisas:

- Falta de información inmediata para la toma de decisiones.
- Ambiente de alta competencia interna, es decir, no se trabaja bajo un esquema con enfoque 100% al cliente o de relaciones como se vio en el capítulo IV.
- Dificultad para obtener reportes.
- Información inconsistente y en múltiples plataformas.
- Problemas de desempeño en los equipos centrales.
- Necesidad de mejorar el desempeño de la empresa.
- Incursión en nuevos mercados
- Implementación de SAP
- Necesidad de predecir el comportamiento de los clientes en cuanto a su uso y capacidad de pago.
- Aumento de las actividades de comercio electrónico.
- Necesidad de adquirir clientes y retener a los ya existentes.
- Necesidad de combatir y prevenir conductas fraudulentas.
- Reducción de costos.

V.2 ANÁLISIS FODA Y ANÁLISIS IEM

El análisis FODA, o análisis de las Fortalezas, Oportunidades, Debilidades y Amenazas, es una de las herramientas esenciales que provee de los insumos necesarios al proceso de planeación estratégica, proporcionando la información necesaria para la implantación de acciones y medidas correctivas y la generación de nuevos o mejores proyectos de mejora.

Las fortalezas y debilidades corresponden al ámbito interno de la organización, y dentro del proceso de planeación estratégica, se debe realizar el análisis de cuáles son esas fortalezas con las que cuenta y cuáles las debilidades que obstaculizan el cumplimiento de sus objetivos estratégicos. ^[1] Por el otro lado, las oportunidades y amenazas, normalmente se relacionan con el ámbito externo de la empresa (clientes, competidores, proveedores).

Realizando este análisis en la operación de RHC se obtiene la siguiente matriz de resultados:

Factores Internos Factores Externos	FORTALEZAS F1. Crecimiento sostenido del 5% de las ventas F2. Producto líder en su ramo F3. Alianzas estratégicas con empresas del ramo F4. Adquisición de nuevos hoteles F5. 70,000 socios cautivos F6. Socios cautivos interesados en el uso F7. Precios competitivos	DEBILIDADES D1. Falta de conocimiento de los socios D2. Perfil de los socios D3. Empresa familiar D4. Documentación de procesos D5. Cultura organizacional D6. Información de varias fuentes inconsistente D7. Gran rotación de personal D8. Análisis reactivos D9. Predicción de comportamiento de socios
OPORTUNIDADES O1. Apoyo al turismo en México O2. Identificación de nuevos segmentos en el mercado O3. Campañas dirigidas a segmentos O4. Esfuerzos mejor enfocados O5. Más empresas interesadas en realizar alianzas	FO (Maxi - Maxi) Estrategia para Maximizar tanto las F como las O	DO (Mini - Maxi) Estrategia para Minimizar las D y Maximizar las O
AMENAZAS A1. Baja del mercado de E.U. en México A2. Nuevos competidores A3. Desastres naturales A4. Inseguridad en México A5. Crecimiento de la deuda con acreedores A6. Fraudes A7. Insuficiencia de los sistemas operativos	FA (Maxi - Mini) Estrategia para Maximizar las F y minimizar las A	DA (Mini - Mini) Estrategia para Minimizar tanto las D como las A

Con base en esta matriz se proponen las siguientes estrategias:

V.2.1 FO (Maxi – Maxi): Utilizar las estrategias de Database Marketing para dirigir las campañas a segmentos internos específicos y así mantener el número de socios cautivos (esfuerzo para invertir más recursos en mantener a los socios actuales en vez de adquirir tantos otros nuevos).

1. Metodología para el Análisis FODA, IPN Dirección de Planeación y Organización, Marzo 2002

V.2.2 DO (Mini – Maxi): Enfocar la cultura organizacional al cliente mediante la documentación de procesos y análisis predictivos con el fin de identificar nuevos segmentos del mercado y realizar un *bench marking* o comparación con otras empresas del ramo que ayude a la mejora de los productos.

V.2.3 FA (Maxi – Mini): Buscar nuevas alianzas con empresas del ramo con el fin de ofrecer una gama más amplia de productos y así hacer frente a la aparición de los nuevos competidores.

V.2.3 DA (Mini – Mini): Tener un solo almacén de datos que provea de la misma información a toda la empresa al igual que aplicaciones vía web que ayuden a la manipulación de datos con el fin de tomar decisiones rápidas y detectar desviaciones o posibles transacciones fraudulentas de manera casi inmediata.

El Modelo Evolutivo de la Información o **análisis IEM** por sus siglas en inglés (*Information Evolution Model*) se basa en el ciclo de vida de la empresa, que se detallaron en el capítulo III.4, enfocando el análisis en el nivel de madurez que tiene la organización para manejar la información como activo estratégico. El análisis IEM maneja cinco niveles que son:

1. **Operar**, es decir, la información se utiliza para que la empresa “sobreviva” y se opera sea como sea.
2. **Consolidar**, la información ayuda a la repetición de acciones y la empresa comienza a operar consistentemente.
3. **Integrar**, se estabiliza el valor de la información y se alcanza la confianza en las tareas.
4. **Optimizar**, se enfocan los esfuerzos en la mejora en la calidad de la información.
5. **Innovar**, se definen nuevas necesidades de información para alcanzar un mayor nivel y calidad operativos.

El análisis IEM se basa en que la empresa debe cumplir ciertas condiciones para que funcione correctamente:

- El cambio debe ser gradual para mejorar.
- Cada nivel debe basarse en el anterior, es decir, no se saltan niveles.
- Los más aptos logran mayores beneficios.
- Requiere de un enfoque proactivo.
- Provee de ventajas competitivas.
- Es constante.
- Es a largo plazo.

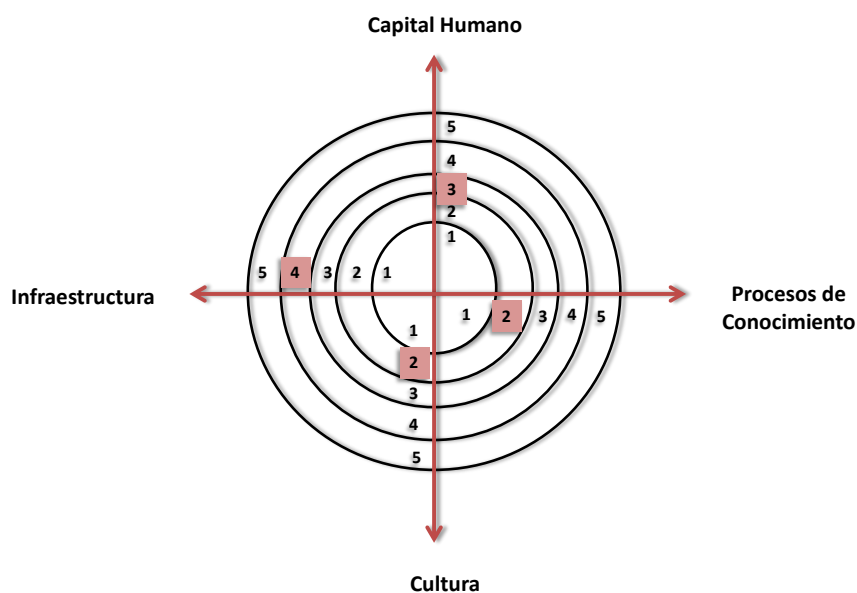
Las habilidades necesarias en una economía global, interconectada y que maneja tecnología de punta, tienen que evolucionar del ámbito técnico al humano. Las organizaciones están automatizando y delegando a proveedores externos todos los procesos lineales para así poder operar en un ambiente de Inteligencia de Negocios. Estas empresas, que basa su operación en datos precisos, accesibles y útiles son las que están despuntando a todos los niveles. La innovación como una competencia medular de la organización reside en la habilidad de ésta

en alinear su infraestructura, procesos, personal y cultura para progresar a través del Modelo Evolutivo de la Información.

Para determinar la madurez de una empresa se deben mapear en los 5 niveles descritos arriba, 4 dimensiones que son:

- Capital humano
- Infraestructura
- Procesos de conocimiento
- Cultura

Con base en este mapeo, el caso para RHC da como resultado la figura V.2:



Esto quiere decir que en cuanto a capital humano, RHC se encuentra en una etapa de integración, estabilizando la permanencia de los colaboradores y las tareas que desempeñan. En cuanto a la cultura organizacional, a pesar de tener casi 30 años en el mercado, se opera con base en repeticiones de las tareas con esfuerzos enfocados en consolidar la información, esto como resultado de que no se ha logrado pasar de la estructura de una empresa familiar a la de un corporativo como tal; la cultura de grupos vs grupos impera en la operación del día a día. Por lo mismo, los procesos de conocimiento se encuentran en el mismo nivel y se dan inconsistencias en la información que hacen que un mismo análisis o estrategia se tenga que repetir varias veces para lograr el resultado esperado; algunos departamentos han desarrollado más el conocimiento y manejo de la información pero no es algo que sea una constante en toda la empresa. Donde se ha avanzado más es en el tema infraestructura puesto que la idea de mejorar los sistemas es una meta constante para la empresa; a pesar de esto, como se vio en los antecedentes, no se ha logrado tener la información consolidada en un solo repositorio y por ende, la información no siempre es confiable y no se pueden tomar decisiones de manera óptima.

V.3 ANÁLISIS DE ESTRATEGIAS Y PROPUESTAS DE MEJORA UTILIZANDO INTELIGENCIA NEGOCIOS

De acuerdo con las conclusiones arrojadas por el análisis FODA el área de Mercadotecnia de RH, diseñó una serie de campañas con el fin de incrementar la penetración de socios y aumentar el número de reservaciones en general y la ocupación en hoteles propios y así aumentar el ingreso de flujos de efectivo a la empresa.

A continuación se presentan las campañas con su descripción, forma de acercamiento a los socios, resultados obtenidos y propuestas de mejora para futuras ocasiones:

V.3.1 CAMPAÑAS DE USO

Nombre de la campaña	USO NUEVA YORK
Fecha de Inicio	16-ene-12
Fecha de Termino	30-ene-12
Duración de la campaña	14 días
Segmento	Socios Mexicanos
Total Envios	7.627
Para Viajar Antes de	27-feb-12
Entregas exitosas	7.576
% Entregas	99,33%
Correos abiertos	3.264
% Correos abiertos	42,80%
Éxito de la campaña	1
% de Éxito	0,01%



Nombre de la campaña	USO RIO OTHON PALACE
Fecha de Inicio	15-ene-12
Fecha de Termino	30-ene-12
Duración de la campaña	15 días
Segmento	Socios Premium Sudamérica
Total Envios	1.635
Para Viajar Antes de	29-abr-12
Entregas exitosas	1.628
% Entregas	99,57%
Correos abiertos	650
% Correos abiertos	39,76%
Éxito de la campaña	3
% de Éxito	0,18%

Ubicado en una zona privilegiada, cercano a todos los atractivos de **Río de Janeiro**, en donde podrás disfrutar de toda la alegría brasileña en las **playas de Copacabana**.

Activa alguno de los certificados de tu membresía **PREMIUM VACATIONS** y **reservando 4 noches** antes del 31 de enero **paga sólo 3***

Llama hoy mismo a tu asesor vacacional desde:

- Argentina (011) 4006 4006
- Chile (02) 450 1060
- Uruguay (02) 901 6302
- Venezuela (212) 278 6595
- Otros países de Sudamérica (04-11) 4006 4006

¡Y disfruta del divertido ambiente carioca!

*Para reservaciones realizadas antes del 31 de enero, válidas del 1 de mayo al 30 de mayo del 2012. Promoción válida únicamente para socios Premium. No aplica para el pago de Casas Rotundas. Copia a disposición de socios en temporada Prime. Sujeto a disponibilidad. No aplica para reservas realizadas a través de

Nombre de la campaña	USO EUROPA
Fecha de Inicio	15-dic-11
Fecha de Terminó	30-dic-11
Duración de la campaña	15 días
Segmento	Socios Mexicanos
Total Envios	8.668
Para Viajar Antes de	30-mar-12
Entregas exitosas	8.631
% Entregas	99,57%
Correos abiertos	3.137
% Correos abiertos	36,19%
Éxito de la campaña	3
% de Éxito	0,03%



Nombre de la campaña	USO HILTON SAN VALENTÍN
Fecha de Inicio	07-feb-12
Fecha de Terminó	14-feb-12
Duración de la campaña	7 días
Segmento	Socios Mexicanos
Total Envios	6.253
Para Reservar Antes de	14-feb-12
Entregas exitosas	6.220
% Entregas	99,47%
Correos abiertos	2.699
% Correos abiertos	43,16%
Éxito de la campaña	13
% de Éxito	0,21%

Este 14 de febrero invita a tu pareja a disfrutar en el Hotel Hilton del centro histórico, tenemos una propuesta increíble para ti.

El paquete incluye:
Habitación decorada con flores y botella de vino espumoso, cena romántica de 3 tiempos para dos personas por solo 5,800* Créditos Vacacionales.

Llama hoy mismo a tu Asesor Vacacional al
5627-8800
¡No pierdas esta oportunidad!

*Incluye: habitación decorada con flores, botella de vino espumoso, cena romántica de 3 tiempos para 2 personas y estacionamiento para un auto. No incluye bebidas durante la cena. Aplica cuota de \$19 USD por complemento de reservación.

Como se puede observar, en todas estas campañas la estrategia fue el envío de un correo electrónico a los socios para ofrecerles la oferta de una determinada temporada. Si bien se utilizó la Minería de Datos para obtener grupos de socios con características específicas (nacionalidad y tipo de membresía determinada, con cuenta de correo electrónico registrada), el envío se hizo a todo un grupo de socios sin una segmentación específica, es decir, no se tomaron en cuenta los hábitos de viaje, el historial de uso de la membresía, la cantidad de ofertas por email que habían sido entregadas a sus direcciones exitosamente y a su vez, a cuáles habían respondido o, en el caso de la campaña de cobranza, el comportamiento de años anteriores en cuanto al pago de la cuota anual.

Por lo anterior, se puede observar que el éxito de las campañas de uso, en ningún caso rebasó el 1%, lo cual es un indicador importante sobre la calidad de las campañas y sobre la necesidad de implementar estrategias de Inteligencia de Negocios y de Mercadotecnia en la empresa. En cuanto a las campañas de cobranza, el éxito fue inferior al 20% lo que trae como consecuencia, esfuerzos posteriores de cobranza telefónica, que aumentarán los gastos.

Por otro lado, a pesar de que los envíos de correos electrónicos masivos, tienen un costo mucho menor en comparación con el envío de correo directo, se está incurriendo en gastos de diseño de imágenes, sueldos, mantenimiento a los servidores,

etc. Que están muy por debajo del beneficio que está trayendo el que 13 socios respondan favorablemente a una determinada campaña. Si bien, la cobranza superior a 1M de USD genera ingresos importantes a la empresa, el flujo de efectivo necesario para hacer frente a las obligaciones de pago a proveedores y préstamos que la empresa tiene presupuestadas, es mucho mayor lo que tendrá consecuencias financieras graves.

Para elaborar propuestas que mejoren las estrategias de campañas de mercadotecnia y así tomar mejores decisiones, se utilizarán de manera integrada el Proceso de la Minería de Datos y el Proceso de Investigación de Mercados descritos en los capítulos II.4 y IV.2, respectivamente.

V.3.1.1 Acuerdo sobre el propósito de la investigación.

V.3.1.1.1 Definición de los problemas:

V.3.1.1.1.1 No se tienen identificados grupos de socios específicos con características en común. Al determinar estos grupos se pueden crear campañas específicas dirigidas a las preferencias de los socios.

V.3.1.1.1.2 No se realizan muestras representativas a las cuales enviar las campañas antes de hacerlo a toda la base de socios. Al elegir una muestra representativa se mide el impacto esperado de lanzar o no una campaña.

V.3.1.1.1.3 No se elaboran predicciones sobre la redención de una campaña determinada, lo que lleva a tener gastos innecesarios. Al elaborar un modelo predictivo se pueden tomar decisiones sobre el envío de la campaña, e incluso, ser preventivos para no lanzar un producto que con anticipación se detecte que no va a tener el éxito que sólo se pensó que tendría.

V.3.1.1.1.4 No se tiene un análisis de retorno de inversión previo lanzamiento para presentar la campaña y tomar decisiones. Al realizar este análisis se podrán justificar de una manera más adecuada los gastos en los que se incurrirán en el lanzamiento de la campaña.

V.3.1.1.2 Alternativas de decisión: Dependiendo de la respuesta de las muestras representativas, el éxito predicho y del retorno de inversión, se tomarán las decisiones de lanzar una campaña de ocupación o cobranza.

V.3.1.1.3 Usuarios de la investigación:

V.3.1.1.3.1 Directores de las áreas de Operaciones y Finanzas

V.3.1.1.3.2 Gerentes de las áreas de Call Center, Mercadotecnia, Administración de Cartera, Tesorería y Presupuestos.

V.3.1.1.3.3 Supervisores de Call Center

V.3.1.1.3.4 Ejecutivos de Call Center, Mercadotecnia y analistas de Administración de Cartera, Tesorería y Presupuestos.

V.3.1.2 Objetivos de la Investigación.

V.3.1.2.1 Crear campañas de mercadotecnia específicas determinando grupos específicos de socios de acuerdo a sus preferencias y atributos como son nacionalidad, ciudad, edad, estado civil.

V.3.1.1.2.2 Elegir una muestra representativa para medir el impacto esperado al lanzar o no una campaña.

V.3.1.1.2.3 Toma de decisiones para enviar o no una campaña, utilizando un modelo predictivo que pueda ayudar a ser preventivos para no lanzar un producto que con anticipación se detecte que no va a tener el éxito que sólo se pensó que tendría.

V.3.1.1.2.4 Justificar de una manera más adecuada los gastos en los que se incurrirán en el lanzamiento de la campaña.

V.3.1.1.2.5 Conseguir que los socios utilicen servicios adicionales al hospedaje, como cenas en restaurantes, uso de spa, etc.

V.3.1.3 Diseño de la investigación

V.3.1.3.1 Aplicando el análisis de relaciones más utilizado en mercadotecnia, se segmentará la base de datos de los socios de acuerdo a las siguientes características:

- Nacionalidad:
 - Socios mexicanos
 - socios de Estados Unidos
 - Socios de Sudamérica
- Nivel de uso en los últimos tres años:
 - Socios con frecuencia de uso mayor a una vez al año
 - Socios con frecuencia de uso igual a una vez al año
 - Socios con frecuencia de uso una vez cada dos años
 - Socios con frecuencia de uso una vez cada tres años
- Preferencias de fechas de viaje:
 - Temporada Alta
 - Temporada Baja – Principios de año
 - Temporada Baja – Otoño
- Email:
 - Socios en email registrado en la base
 - Socios que no cuentan con email registrado en la base

Nacionalidad		Frecuencia	Preferencias Viajes		Email		
Mexicanos	32.678,00	+1 vez al año	1.634	Temporada Alta	1.062	Sí	297
						No	765
				Temporada Baja Sept - Noviembre	359	Sí	101
				No	259		
				Temporada Baja Principios de año	212	Sí	59
						No	153
		1 vez al año	4.902	Temporada Alta	3.186	Sí	892
						No	2.294
				Temporada Baja Sept - Noviembre	1.078	Sí	302
				No	776		
				Temporada Baja Principios de año	637	Sí	178
						No	459
		1 vez c/ 2 años	11.111	Temporada Alta	7.222	Sí	2.022
						No	5.200
				Temporada Baja Sept - Noviembre	2.444	Sí	684
				No	1.760		
		Temporada Baja Principios de año	1.444	Sí	404		
				No	1.040		
1 vez c/3 años	6.536	Temporada Alta	4.248	Sí	1.189		
				No	3.059		
		Temporada Baja Sept - Noviembre	1.438	Sí	403		
		No	1.035				
		Temporada Baja Principios de año	850	Sí	238		
				No	612		
		No han viajada en 3 años	8.496	N/A	N/A		
USA	25.219,00	+1 vez al año	2.018	Temporada Alta	948	Sí	303
						No	645
				Temporada Baja Sept - Noviembre	767	Sí	245
				No	521		
				Temporada Baja Principios de año	303	Sí	97
						No	206
		1 vez al año	7.061	Temporada Alta	3.319	Sí	1.062
						No	2.257
				Temporada Baja Sept - Noviembre	2.683	Sí	859
				No	1.825		
				Temporada Baja Principios de año	1.059	Sí	339
						No	720
		1 vez c/ 2 años	7.818	Temporada Alta	3.674	Sí	1.176
						No	2.499
				Temporada Baja Sept - Noviembre	2.971	Sí	951
				No	2.020		
		Temporada Baja Principios de año	1.173	Sí	375		
				No	797		
1 vez c/3 años	3.783	Temporada Alta	1.778	Sí	569		
				No	1.209		
		Temporada Baja Sept - Noviembre	1.437	Sí	460		
		No	977				
		Temporada Baja Principios de año	567	Sí	182		
				No	386		
		No han viajada en 3 años	4.539	N/A	N/A		
Sudamérica	14.648,00	+1 vez al año	1.611	Temporada Alta	790	Sí	95
						No	695
				Temporada Baja Sept - Noviembre	419	Sí	50
				No	369		
				Temporada Baja Principios de año	403	Sí	48
						No	354
		1 vez al año	3.223	Temporada Alta	1.579	Sí	189
						No	1.390
				Temporada Baja Sept - Noviembre	838	Sí	101
				No	737		
				Temporada Baja Principios de año	806	Sí	97
						No	709
		1 vez c/ 2 años	2.637	Temporada Alta	1.292	Sí	155
						No	1.137
				Temporada Baja Sept - Noviembre	686	Sí	82
				No	603		
		Temporada Baja Principios de año	659	Sí	79		
				No	580		
1 vez c/3 años	2.344	Temporada Alta	1.148	Sí	138		
				No	1.011		
		Temporada Baja Sept - Noviembre	609	Sí	73		
		No	536				
		Temporada Baja Principios de año	586	Sí	70		
				No	516		
		No han viajada en 3 años	4.834	N/A	N/A		
TOTAL							72.545,00

Tabla V.3.1.3 Segmentación de la base de socios

Con base en los resultados arrojados por la segmentación y debido a que el objetivo es incrementar el flujo de efectivo utilizando los espacios excedentes en fechas que no son temporada alta, se toma la decisión de dirigir las campañas hacia los segmentos sombreados, que son los más propensos a responder positivamente al tipo de campaña que se pretende lanzar. No se tomarán para este ejercicio, los socios que viajan más de una vez al año puesto que son personas más propensas a realizar un viaje aun cuando no haya una campaña especial para ellos.

Así, los 12 segmentos elegidos son:

- a) Socios mexicanos que viajan una vez al año en temporada baja Septiembre-Noviembre
- b) Socios mexicanos que viajan una vez al año en temporada baja principios de año
- c) Socios mexicanos que viajan una vez cada dos años en temporada baja Septiembre-Noviembre
- d) Socios mexicanos que viajan una vez cada dos años en temporada baja principios de año
- e) Socios de E.U. que viajan una vez al año en temporada baja Septiembre-Noviembre
- f) Socios de E.U. que viajan una vez al año en temporada baja principios de año
- g) Socios de E.U. que viajan una vez cada dos años en temporada baja Septiembre-Noviembre
- h) Socios de E.U. que viajan una vez cada dos años en temporada baja principios de año
- i) Socios de SUDAMÉRICA que viajan una vez al año en temporada baja Septiembre-Noviembre
- j) Socios de SUDAMÉRICA que viajan una vez al año en temporada baja principios de año
- k) Socios de SUDAMÉRICA que viajan una vez cada dos años en temporada baja Septiembre-Noviembre
- l) Socios de SUDAMÉRICA que viajan una vez cada dos años en temporada baja principios de año

V.3.1.4 Preparación y Análisis de los datos

Se elegirá una muestra representativa de cada uno de los segmentos seleccionados a la cual anunciar la campaña vía correo electrónico y por teléfono, antes de enviar una pieza por correo directo.

Ya que es la primera vez que se realizará un muestreo aleatorio, se asumirán los siguientes parámetros:

- El margen de error permitido será de $d = \pm 5\%$
- El riesgo de que el error sea mayor de ese margen permitido será de $\alpha = \frac{1}{20}$

Utilizando la fórmula para obtener n :

$$n = \frac{n_0}{1 + (n_0 - 1)/N}$$

Donde $n_0 = \frac{t^2 pq}{d^2}$, $N =$ Tamaño de socios por segmento,

$t =$ Abscisa de la curva normal que corta un área de tamaño α en los extremos o colas.

Entonces

$$n_0 = \frac{(4)(0.5)(0.5)}{(0.0025)} = 400$$

Por lo que, para cada segmento, el tamaño de la muestra queda como sigue:

- 292 socios mexicanos, un viaje al año, temporada baja sept-nov.
- 246 socios mexicanos, un viaje al año, temporada baja principios
- 344 socios mexicanos, un viaje cada 2 años, temp. Baja sept – nov.
- 313 socios mexicanos, un viaje cada 2 años, temp. Baja principios
- 348 socios E.U., un viaje al año, temporada baja sept – nov.
- 291 socios E.U., un viaje al año, temporada baja principios
- 353 socios E.U., un viaje cada 2 años, temp. Baja sept – nov.
- 298 socios E.U., un viaje cada 2 años, temp. Baja principios
- 271 socios Sudamérica, un viaje al año, temporada baja sept - nov.
- 268 socios Sudamérica, un viaje al año, temporada baja principios
- 253 socios Sudamérica, un viaje cada 2 años, temp. Baja sept – nov
- 249 socios Sudamérica, un viaje cada 2 años, temp. Baja principios

Si se toma en cuenta los resultados de redención obtenida en las campañas en las que sólo se utilizó el email se tendrían que, para estas muestras representativas, el número de socios que realizarían una reservación y su consiguiente ingreso esperado sería:

Segmento	Tamaño	Ingreso (USD)
a	292	\$21,74
b	246	\$18,31
c	344	\$25,60
d	313	\$23,33
e	348	\$25,92
f	291	\$21,63
g	353	\$26,25
h	298	\$22,22
i	271	\$155,12
j	268	\$153,14
k	253	\$144,74
l	249	\$142,65

Considerando que por cada reserva que se haga se obtendrá un ingreso de 312 usd.

Por lo que un esfuerzo de llamadas es forzoso para incrementar el contacto con el cliente y posicionar la campaña.

Nuevamente haciendo uso de la información almacenada en el data mart de reservaciones y uso de la empresa, tenemos que, históricamente, el incremento en reservaciones desde 2007 a la fecha, utilizando el esfuerzo de llamadas se ha dado como sigue:

Año	Reservaciones orgánicas	Reservas con análisis de socios	Diferencia	%
2007	57.375	57.662	287	0,50%
2008	60.679	60.982	303	0,50%
2009	63.918	64.302	384	0,60%
2010	67.151	67.622	470	0,70%
2011	70.378	70.941	563	0,80%
2012	73.891	74.261	369	0,50%
PROMEDIO				0,60%

Se puede entonces esperar un incremento del 0.60% para las reservaciones derivadas de las campañas a lanzar, debido además, a que los segmentos a los que se dirigen son segmentos propicios a realizar viajes, entonces se tendría que:

Segmento	Tamaño	Ingreso (USD)	Ingreso con llamadas	Incremento
a	292	\$21,74	\$546,57	\$524,83
b	246	\$18,31	\$460,47	\$442,16
c	344	\$25,60	\$643,72	\$618,12
d	313	\$23,33	\$586,72	\$563,39
e	348	\$25,92	\$651,87	\$625,95
f	291	\$21,63	\$543,91	\$522,28
g	353	\$26,25	\$660,14	\$633,89
h	298	\$22,22	\$558,70	\$536,49
i	271	\$155,12	\$507,24	\$352,12
j	268	\$153,14	\$500,78	\$347,64
k	253	\$144,74	\$473,31	\$328,57
l	249	\$142,65	\$466,45	\$323,80

Es claro que el incremento en los ingresos sería considerable si se realiza un plan de llamadas para cada uno de los segmentos elegidos.

Entonces se toma la decisión de realizar una llamada a cada socio de la muestra, para ofrecer la promoción y además enviar un email a las direcciones de correo que se encuentren en la base.

V.3.1.5 Resultados de la Investigación y toma de decisiones

Después de llevar a cabo las acciones sobre las muestras, considerando un costo de 3 usd por llamada realizada y 0.50 centavos por email enviado, los resultados fueron los siguientes:

Segmento	Tamaño	Redención	Socios	Ingreso	Costos	Rendimiento
a	292	13%	38	\$11.842,32	\$890,51	\$10.951,81
b	246	2%	5	\$1.534,91	\$750,24	\$784,67
c	344	9%	31	\$9.655,82	\$1.048,80	\$8.607,02
d	313	0%	0	\$0,00	\$955,93	-\$955,93
e	348	19%	66	\$20.642,51	\$1.062,07	\$19.580,44
f	291	22%	64	\$19.943,34	\$886,18	\$19.057,16
g	353	12%	42	\$13.202,77	\$1.075,55	\$12.127,22
h	298	1%	3	\$931,17	\$910,28	\$20,89
i	271	20%	54	\$16.908,16	\$826,44	\$16.081,72
j	268	25%	67	\$20.865,96	\$815,91	\$20.050,05
k	253	3%	8	\$2.366,57	\$771,16	\$1.595,41
l	249	11%	27	\$8.551,59	\$759,98	\$7.791,61

De lo anterior podemos deducir, que hay segmentos de socios a los que las campañas diseñadas no deberían lanzarse, sobre todo si, adicional a los esfuerzos de llamadas y correo electrónico, se quiere enviar una pieza de correo directo, lo que implicará más gasto de diseño, manufactura y envío.

Se toma entonces la decisión de eliminar los segmentos cuyo rendimiento es menor a 10,000 usd, esto porque, además, se tienen que considerar los sueldos de los ejecutivos y personal involucrado en el contacto con los socios.

De tal suerte que, el retorno de inversión (ROI) esperado para las campañas a, e, f, g, i y j, tomando en cuenta el tamaño total del segmento y la redención de la muestra, será de:

$$ROI = \frac{\text{beneficio obtenido} - \text{inversión}}{\text{inversión}}$$

Segmento	Tamaño	Redención	Socios	Ingreso	Costos	Rendimiento	ROI
a	1078	13%	140	\$43.738,85	\$3.289,04	\$40.449,81	12,30
e	2683	19%	510	\$159.066,12	\$8.184,07	\$150.882,05	18,44
f	1059	22%	233	\$72.703,35	\$3.230,55	\$69.472,80	21,50
g	2971	12%	356	\$111.226,68	\$9.060,93	\$102.165,75	11,28
i	838	20%	168	\$52.282,81	\$2.555,49	\$49.727,32	19,46
j	806	25%	201	\$62.839,92	\$2.457,20	\$60.382,72	24,57

Con estos resultados se puede fácilmente justificar el lanzamiento de las campañas con las bases aportadas por la minería de datos, la investigación de mercados y la toma de decisiones.

Se puede proceder de distintas maneras y utilizar diversas herramientas y procesos para anticipar resultados y evitar pérdidas y esfuerzos innecesarios.

Aún más, el análisis de enlaces o conexiones (Market Basket) sería una herramienta importante, ya que se podrían ofrecer servicios adicionales al hospedaje, como el Todo Incluido o cenas en restaurantes, uso de spa, etc. Consultando las preferencias históricas que han mostrado los socios a lo largo de su permanencia dentro del almacén de datos de RH.

Esto traería como resultado que se sumaran a los ingresos recibidos por reservaciones, los pagos de dichos servicios utilizando las mismas vías, es decir, no habría costos adicionales en cuando a la promoción.

En resumen, los resultados que una empresa que sólo utiliza parcialmente la Inteligencia de Negocios para su operación, son claros en este ejemplo, ya que no se está obteniendo el éxito deseado (más no esperado, porque no hay análisis que lo demuestre), lo que trae pérdidas económicas y mucho peor, socios descontentos o que simplemente, no están siendo satisfechos con los nuevos productos o estrategias que se están proponiendo. En cambio, cuando se utilizan algunas técnicas de Minería de Datos e Investigación de Mercados se pueden tomar decisiones con tiempo y conocimiento de que los resultados esperados, serán favorables para la empresa.

CONCLUSIONES

Conforme a lo desarrollado anteriormente, se puede determinar qué factores son indispensables para una operación organizacional basada en la Inteligencia de Negocios:

- Un robusto almacén central de datos en donde se encuentre toda la información correspondiente a los clientes (datos personales, transaccionales, demográficos).
- Un equipo enfocado al mantenimiento y control de dicho almacén.
- Un área de Minería de Datos, encargada de la explotación de la información contenida en el almacén de datos y de desarrollar herramientas que sirvan para la manipulación inmediata de la misma con el fin de obtener de manera casi inmediata, los conocimientos necesarios que las áreas de la empresa, desde los directivos hasta las área de operaciones, necesitan.
- Un área de Database Marketing, especializada en el análisis de información, con los conocimientos matemáticos y estadísticos necesarios, que puedan canalizar las peticiones de las distintas áreas que tienen contacto con los clientes, y que indiquen con fundamentos y resultados demostrados, cuándo dicha información generará un valor agregado para la empresa y cuándo no.
- Un área de Inteligencia de Mercadotecnia encargada en generar campañas enfocadas, específicas y bien fundamentadas.
- Generar una cultura organizacional en el que los esfuerzos de todas las áreas estén enfocados al cliente como la pieza más importante del negocio.
- Contar con un software especializado que ayude a que la consistencia de la información y a que sólo exista una plataforma para la consulta y la explotación de los datos.
- Encargados de la toma de decisiones que estén preparados y capacitados para convencer a los equipos de trabajo y no sólo impongan sus ideas.

Con todas estas piezas claves dentro de un proceso, con responsabilidades bien definidas e información confiable, se logrará cumplir con los objetivos de reducción de costos y sobre todo, maximización de las ganancias. De igual manera ayudará a la retención de los colaboradores aumentando la efectividad y eficiencia de los mismos, creando así un equipo dinámico, en un ambiente en el que la comunicación, la flexibilidad de horarios y los programas de motivación funcionen de manera constante. Las habilidades y experiencia de todos los miembros de la empresa es un punto importante a considerar cuando se lleva a cabo un proyecto de inteligencia de negocio, no se debe minimizar las habilidades de gente que lleva años en una empresa ni tampoco la experiencia de gente joven que recién se integra: algunas organizaciones, por ejemplo, pretenden que los administradores de los datos, que típicamente son analistas de negocios con muy poca experiencia técnica diseñen las bases de datos multidimensionales. Se debe contar con gente especializada en cada tarea que aporte sus conocimientos a todos los eslabones de la cadena, para esto, es muy necesario que la empresa, en muchas ocasiones invierta en el entrenamiento de sus integrantes.

Una empresa con colaboradores contentos generará clientes contentos y ese debe ser el principal objetivo de una organización. Cuando se pierde de vista dicho objetivo y cada área sólo trabaja en función de cumplir las tareas que se le han asignado, es cuando comienza a crearse un ambiente de competitividad en el que lo importante ya no es la información o los resultados que se generen y que puedan servir para que otra área arme sus estrategias sino en cómo defender, muchas veces sin fundamentos la información que se está generando como la única información valiosa de la empresa.

Un ambiente de Inteligencia de Negocios ayuda a que todo este tipo de malas prácticas se minimicen, si no es que desaparezcan del trabajo del día a día y que se puedan también diversificar esfuerzos en establecer controles y medidas que puedan ayudar a prevenir desde transacciones fraudulentas como a pérdidas importantes de capital.

BIBLIOGRAFÍA

1. Business Intelligence: Data Mining and Optimization for Decision Making
Carlo Vercellis
Wiley and Sons Ltd
2009
2. Data Mining Explained: A Manager's Guide To Customer-Centric Business Intelligence
Rhonda Delmater, Monte Hancock
Digital Press
2001
3. Data Mining Concepts And Techniques
Jiawei Han, Micheline Kamber
Academic Press, Morgan Kaufmann publishers
2001
4. Knowledge, Discovery And Data Mining: Challenges And Realities
Xingquan Zhu, Ian Davidson
Premier Reference Source
2007
5. Managerial Decision Making: A Guide To Successful Business Decisions
Alan J. Rowe, James D. Boulgarides
Prentice Hall, Inc.
1994
6. La Toma De Decisiones: Harvard Business Review
P. Drucker, J. Hammond, R. Keeney, H. Raiffa, A. Etzioni, C. Argyris, P. Stryker, A. Hayashi
Ediciones Deusto
2002
7. Marketing Research
David A. Aaker, V. Kumar, George S. Day
Limusa – Wiley
2001
8. Marketing Research: An Applied Orientation
Naresh K. Malhotra
Prentice Hall
1999
9. Metodología para el Análisis FODA
Dirección de Planeación y Organización, IPN
2002
10. Sampling techniques
Cochran, William G.
John Wiley & Sons, Inc
1977

11. www.businessintelligence.com

Artículos:

- a. An Overview of Business Intelligence Technology
Surajit Chaudhuri, Umeshwar Dayal, Vivek Narasayya
2005
- b. 7 Simple Rules For Successful Real-Time Business Intelligence Implementation
Jaymin Patel
2005
- c. Techniques for Integrating BI Into The Enterprise – Parte 1 y 2
Ferguson Mike
2004
- d. Organizational Barriers to Business Intelligence (Part 1)
Moss Larissa T.
2004