

IN627

Investigación de Mercados

Profesores:
Rodrigo Niño
Emilio Polit

1

Capítulo 12: Nuevas Técnicas de Recolección y Análisis de Datos

2

Nuevas Técnicas de Recolección de Datos

3

Motivación

- En variadas ocasiones debemos recoger información de consumidores de difícil acceso.
- La falta de tiempo y la falta de disposición de este tipo de consumidor para participar en focus group, entrevistas individuales o responder cuestionarios administrados por métodos tradicionales (telefónicos o cara a cara) ha llevado a los investigadores de mercado a buscar nuevas técnicas de recolección de datos.
- La evolución y masificación de las tecnologías de información han permitido explorar nuevos mecanismos de recolección de información.

4

Motivación

- Consumidores de difícil acceso:
 - Jóvenes
 - ⇒ Algunos jóvenes no "están ni ahí" con asistir a focus group o responder cuestionarios extensos cara a cara.
 - Mujeres jefas de hogar
 - ⇒ Todo el tiempo que no están trabajando lo destinan a labores domésticas por lo que se manifiestan imposibilidades de contestar cuestionarios o asistir a entrevistas en profundidad.
 - Profesionales de altos cargos ejecutivos (gerentes, empresarios)
 - ⇒ En su lugar de trabajo no tienen tiempo para recibir un encuestador o contestar una encuesta telefónica ya que su agenda está siempre ocupada.
 - ⇒ Pasan muy poco tiempo en sus casas, no tienen hora fija de llegada a casa, viven en lugares de difícil acceso y cuentan con números telefónicos privados.

5

Nuevas Técnicas de Recolección de Datos

- Existen nuevas técnicas de recolección de datos tanto cualitativos como cuantitativos:
- Técnicas de investigación cualitativa:
 - Etnografía
 - Comprador misterioso
 - Focus group por Internet (Chat group)
 - Entrevistas en profundidad por Internet
- Técnicas de investigación cuantitativa:
 - Geo-referenciación
 - CATI
 - CAPI
 - Encuestas por Internet

6

Nuevas Técnicas de Investigación Cualitativa

Etnografía

- Metodología de investigación cualitativa en que el investigador comparte por un tiempo determinado la vida (hábitos, estilo de vida) del individuo estudiado.
- Como es un tipo de investigación en que el investigador comparte la vida con el individuo y no sólo lo observa, se generan lazos de confianza que permiten acceder a grupos muy difíciles de estudiar con métodos tradicionales de investigación (entrevistas en profundidad u observación).
 - Ejemplos:
 - ⇒ Sectas (barras bravas, religiosas, musicales)
 - ⇒ Poblaciones marginales urbanas y rurales
 - ⇒ Jefes de hogar y dueñas de casa con escasa instrucción (estratos D y E de la población)

7

Nuevas Técnicas de Investigación Cualitativa

Comprador misterioso

- Técnica cualitativa de medición de la calidad de servicio. Permite estudiar la brecha entre el servicio especificado por la empresa y el servicio entregado.
- Se usa un comprador anónimo (espía), el cual registra su interacción con el personal a cargo de la entrega del servicio.
 - Para este registro se usan cámaras ocultas:
 - ⇒ Lápiz cámara
 - ⇒ Reloj cámara
 - ⇒ Anteojos cámara
- El uso de esta técnica de recolección de información requiere de un cuestionamiento ético de parte de los investigadores.

8

Nuevas Técnicas de Investigación Cualitativa

Focus group por Internet:

- Grupos focales en los cuales los participantes interactúan (chatean) por Internet.
- La idea es reunir en una conversación virtual (MSN Messenger) a un pequeño grupo de personas, bajo la supervisión de uno o más moderadores, para que discutan de un tema que se desea investigar en profundidad.

9

Nuevas Técnicas de Investigación Cualitativa

Focus group por Internet:

- **Ventajas:**
 - Permite acceder a grupos que son reacios que a asistir a focus group tradicionales por problemas de disposición o tiempo.
 - ⇒ Ejemplo: profesionales jóvenes.
 - Más barato y menos complejo que los focus group tradicionales ya que se evita juntar a un grupo de personas en un lugar adecuado para desarrollo de la investigación.
 - Permite incorporar fácilmente estímulos audiovisuales (videos, imágenes).
 - Algunos participantes fáciles de inhibirse frente a otras personas tienden a intervenir y opinar más.
 - Para ciertos grupos de la población (ej.:adolescentes ABC1) chatear es un lenguaje de comunicación totalmente natural.

10

Nuevas Técnicas de Investigación Cualitativa

Focus group por Internet:

- **Desventajas:**
 - Se pierde la información que proviene de las reacciones físicas (gestos) de los participantes (se puede solucionar con el uso de cámaras Web).
 - Se pierde la interacción social natural (cara a cara) de las personas.
 - Para ciertos grupos de la población (ej.:dueñas de casa mayores de 45 años) no es natural chatear.
 - En Chile, Internet aún no es masivo es ciertos grupos de la población (C3, D, E) por lo que es difícil acceder a estos grupos con estas metodologías.

11

Nuevas Técnicas de Investigación Cuantitativa

Geo-referenciación

- Metodología de segmentación que cruza datos geográficos (ciudad, comuna, manzana, dirección, número telefónico) con datos socio-demográficos (GSE, sexo, edad, estado civil).
- De esta manera, se obtiene un mapa de la zona geográfica que se desea estudiar (ciudad, comuna) en que cada individuo está identificado.
- De esta manera se puede:
 - Segmentar a priori.
 - Seleccionar targets analizando el mapa de composición geográfica.
 - Fijar precios según la composición geográfica y socio-demográfica de la población.
 - Optimizar la entrega de publicidad a domicilio.
 - Optimizar la publicidad en puntos geográficos.

12

Nuevas Técnicas de Investigación Cuantitativa

CATI

- Encuestas telefónicas asistidas por computador.
- Se tiene la integración del teclado del teléfono con la base de datos (matriz casos-preguntas).

CAPI

- Encuestas personales (cara a cara) asistidas por computador.
- El encuestador debe recoger la información a través de un notebook o una palm.
- Se tiene la integración del teclado del equipo computacional con la base de datos (matriz casos-preguntas).

13

Nuevas Técnicas de Investigación Cuantitativa

- El CATI y el CAPI permiten:
 - Evitar errores de digitación.
 - Evitar errores de administración de cuestionario (ej.: saltos de preguntas).
 - Más rapidez que la encuesta telefónica o cara a cara tradicional.
 - Se debe incurrir en algunos costos de inversión los cuales se recuperan en el mediano plazo por la mayor eficiencia del sistema.
 - Costos de inversión:
 - ⇒ *Capacitación encuestadores*
 - ⇒ *Equipos*

14

Nuevas Técnicas de Investigación Cuantitativa

Encuestas por Internet

- Para ver cómo funcionan y las enormes ventajas de las encuestas autoadministradas vía Internet veamos un exitoso caso de aplicación:

BBDO Research

15

BBDO Research: Encuestas por Internet

Octubre del 2004

16

BBDO Research: Encuestas por Internet

BBDO Research es el área especializada en Investigación de Mercado del Departamento de Planificación Estratégica de BBDO Chile.

Surge de la necesidad de obtener información relevante y útil de los consumidores de una forma rápida y oportuna para la realización de las estrategias comerciales de nuestras marcas. Los estudios de BBDO Research han permitido tener información del consumidor incluso en 24 horas.

Los estudios de BBDO Research han permitido generar un gran ahorro de costos para BBDO Chile por concepto de compra de estudios externos (un estudio de mercado "tipo" cuesta alrededor de \$2.000.000)

17

BBDO Research: Encuestas por Internet

Adicionalmente, BBDO Research genera ingresos propios a través de la venta de estudios a clientes internos y externos.

En octubre del año 2003 nace el proyecto BBDO Research. En noviembre del año 2003 se hace la primera encuesta por Internet.

En febrero del año 2004 se vende la primera encuesta a un cliente interno de BBDO Chile. En abril del año 2004 se vende la primera encuesta a un cliente externo a BBDO Chile.

18

Base de Datos

Actualmente la base de datos de BBDO Chile está compuesta aproximadamente por 8000 personas de la ciudad de Santiago de Chile.

Estas personas han sido ingresadas a esta base de datos respondiendo una encuesta de datos demográficos, la cual es realizada cara a cara por los encuestadores de BBDO Chile en distintos puntos geográficos de gran afluencia.

En los próximos meses esta base se extenderá a otras ciudades del país.

19

Clasificación Sociodemográfica

El único requisito para ser miembro del panel BBDO Chile es tener una dirección de correo electrónico donde recibir encuestas y tener 15 años o más.

La encuesta de datos demográficos recoge la siguiente información de cada entrevistado:

- | | |
|-----------------------------------|---|
| • Dirección de correo electrónico | • Nivel educacional de su jefe de hogar |
| • Nombre | • Nivel ocupacional de su jefe de hogar |
| • Edad | • Tenencia de bienes |
| • Sexo | • Ingreso familiar |
| • Estado civil | |
| • Comuna de residencia | |

20

Clasificación Sociodemográfica

Los integrantes del panel BBDO Chile están clasificados según su grupo socioeconómico (GSE), sexo y edad.

La segmentación por GSE (ABC1,C2 y C3) se realiza de la siguiente forma:

- Un 70% se calcula a través del sistema ESOMAR (European Society for Opinion and Marketing Research) el cual incorpora como variables de clasificación el nivel educacional alcanzado y la categoría ocupacional del jefe del hogar.
- El otro 30% del modelo considera las variables: ingreso familiar (10%), tenencia de bienes (10%) y comuna de residencia (10%).

21

Clasificación Sociodemográfica

Se excluyen del panel BBDO Chile personas de los grupos D y E (bajo) debido al difícil acceso que en Chile tienen a Internet.

22

Sistema de Incentivos

Cada encuesta tiene asociado un puntaje. Cuando un integrante del panel BBDO Chile contesta una encuesta acumula ese puntaje.

Los puntos son acumulables y canjeables por variados premios como:

- Calculadoras
- Relojes
- Radios personales
- Calentadores de agua
- Reproductores DVD
- Equipos de sonido
- Televisores
- Cámaras de video

23

Sistema de Incentivos

BBDO Chile cuenta con el sitio Web www.bbdo.cl/encuestas donde cada miembro del panel puede:

- Revisar su puntaje
- Canjear premios
- Modificar sus datos personales

BBDO Chile ha entregado en un año más de 130 premios a los integrantes de su panel de encuestados.

24

Software Diseño Cuestionario

BBDO Research diseña sus encuestas con el software Survey Crafter el cual permite construir preguntas con diferentes estructuras:

- **Abiertas**
- **Cerradas simples**
- **Cerradas múltiples**
- **Matriciales**
- **Ranking**
- **Cantidad**

25

Software Diseño Cuestionario

Survey Crafter permite insertar en el cuestionario publicidad gráfica y, a través de la inserción de un link a una página Web, también publicidad audiovisual.

Survey Crafter permite que el cuestionario pueda llevarse a un formato Web.

- Ejemplo: www.encuestasbbdo.cl/151004le/encuesta.htm

26

Software Análisis de Datos

BBDO Research analiza los datos que provienen de sus encuestas con el software SPSS.

SPSS permite obtener :

- Estadísticos descriptivos (frecuencias, porcentajes, medias)
- Segmentación de casos y/o variables
- Reducción de variables en factores relevantes
- Mapas de posicionamiento
- Regresiones

27

Ventajas BBDO Research

Multimedia:

Una de las grandes potencialidades del sistema consiste en la **posibilidad de insertar elementos gráficos y audiovisuales** en las encuestas. De esta manera, es factible testear y evaluar campañas publicitarias.

Rapidez:

Es posible obtener muestras de **300-400 casos en sólo 3 días**, período que disminuye a medida que aumenta el número de integrantes del panel BBDO Chile. Esto se vincula al carácter instantáneo que tiene la comunicación vía Internet.

28

Ventajas BBDO Research

Representatividad:

El panel está clasificado a priori por GSE, sexo y edad, lo que permite resguardar la representatividad de la muestra de la población estudiada. Adicionalmente es **factible realizar estudios a targets específicos y de alta selectividad**, sin necesidad de hacer excesivos sobre-muestreos.

Acceso y Flexibilidad:

El privilegiado acceso a población antes inaccesible, como personas con poco tiempo, cargos gerenciales o reacias a responder encuestas realizadas por métodos tradicionales (cara a cara, telefónicas), entrega la **posibilidad de investigar a tipos de consumidores de alcance más complejo**.

29

Ventajas BBDO Research

Menor estructura de costos:

Al realizar una encuesta vía Internet **no es necesario incurrir en elevados costos** de espacio físico, encuestadores, papeleo y digitación.

Tecnología:

El sistema cuenta con la **integración del software de recolección de datos con el software que realiza el análisis estadístico** de los mismos, lo que permite obtener resultados en poco tiempo.

30

Ventajas BBDO Research

Lo anterior convierte a BBDO Research en un sistema confiable y eficiente para:

- Evaluar campañas publicitarias
- Realizar estudios de imagen y posicionamiento
- Tests de conceptos
- Estudios de opinión
- Investigaciones de mercado en general

Todo esto, con una **altísima relación calidad/precio y en menos de una semana.**

31

Ejemplos

32

Evaluación Campaña de la Leche “Yo tomo”

Junio 2004

33

Nota Metodológica

Tipo de estudio: cuantitativo mediante encuestas autoadministradas vía Internet.

Características especiales del estudio: los encuestados presenciaron 2 de los spots de la campaña después de responder la pregunta de recordación espontánea: “¿Qué comerciales de televisión recuerda haber visto en las últimas dos semanas?”.

Grupo objetivo: hombres y mujeres mayores de 15 años de niveles socioeconómicos alto (ABC1) y medio (C2 y C3).

Muestra: 529 casos.

Fecha: 17 al 20 de junio del 2004.

Ponderación: resultados ponderados según distribución por GSE, edad y sexo de la población de la Región Metropolitana.

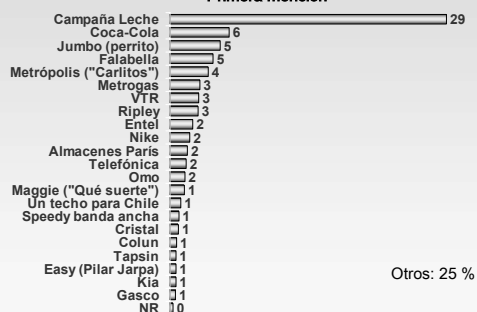
34

¿Qué comerciales recuerda haber visto en televisión en las últimas dos semanas?

Respuesta espontánea

Base: total muestra= 529

Primera mención



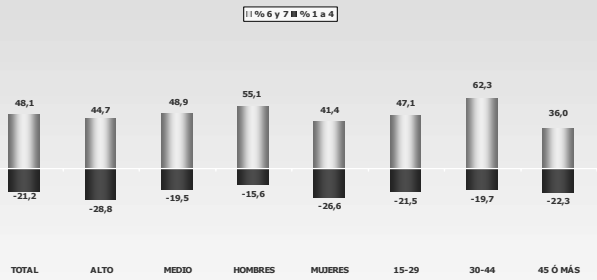
35

Exposición Comerciales Campaña de la Leche “Yo tomo”

36

Nivel de agrado de los comerciales

Escala de 1 a 7, 1=no le agradaron nada, 7=le agradaron mucho
Base: total muestra=529

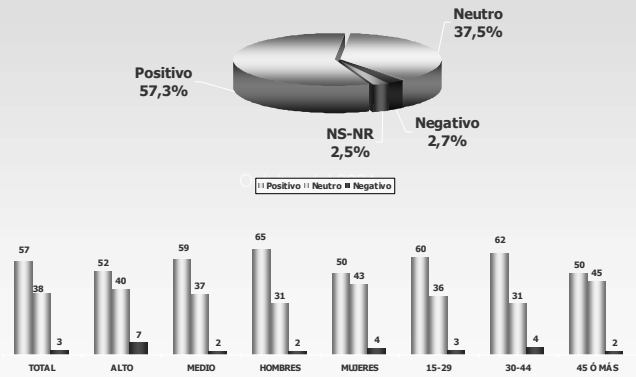


Nota promedio: 5,4

37

Ud. diría que esta campaña tiene un efecto para los chilenos...

Base: total muestra=529



38

Imagen SONY

Agosto 2004

39

Nota Metodológica

Tipo de estudio: cuantitativo mediante encuestas autoadministradas vía Internet.

Grupo objetivo: hombres y mujeres, mayores de 15 años, de niveles socioeconómicos ABC1, C2 y C3.

Muestra: 474 casos.

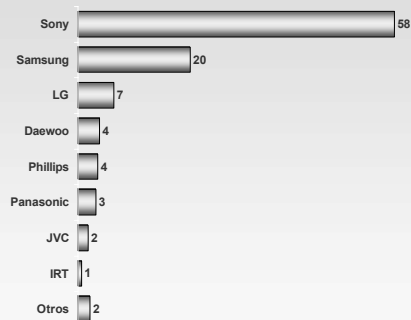
Fecha: 5 - 10 de julio del 2004.

Ponderación: resultados ponderados según distribución por GSE, sexo y edad de la población de la Región Metropolitana.

40

Menciona las marcas de televisores que se te vengan a la mente

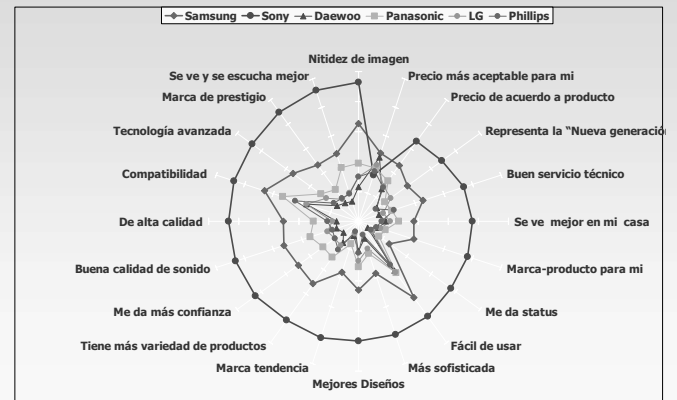
Respuesta espontánea
Base: total muestra = 474



41

Asociación de atributos con empresas

Base: total muestra = 474



42

Evaluación “Agrosuper alimenta”

Julio 2004

43

Nota Metodológica

Tipo de estudio: cuantitativo mediante encuestas autoadministradas vía Internet.

Grupo objetivo: hombres y mujeres, mayores de 15 años, de niveles socioeconómicos ABC1, C2 y C3.

Muestra: 419 casos.

Fecha: 16 - 19 de julio del 2004.

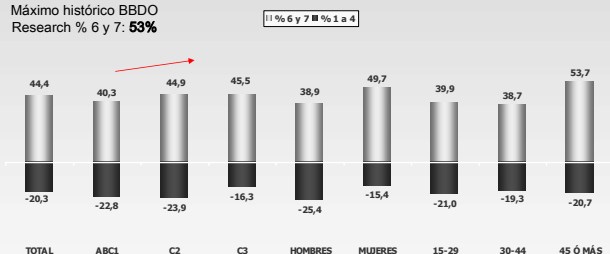
Ponderación: resultados ponderados según distribución por GSE, sexo y edad de la población de la Región Metropolitana.

44

Evaluación slogan “Agrosuper alimenta”

Escala de 1 a 7, 1 = No te gusta nada, 7 = Te gusta mucho
Base: total muestra = 417

Máximo histórico BBDO
Research % 6 y 7: **53%**



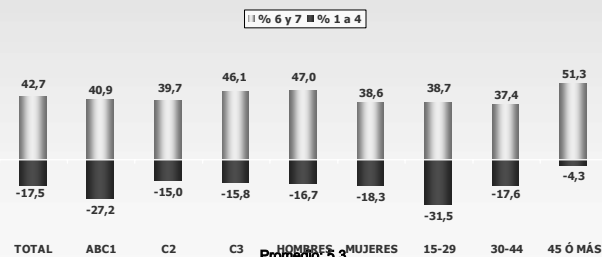
Promedio: 5,3

45

Evaluación logo de Agrosuper

Escala de 1 a 7, 1 = No te gusta nada, 7 = Te gusta mucho
Base: total muestra = 417

Máximo histórico BBDO
Research % 6 y 7: **56%**



Promedio: 5,3

46

Nuevas Técnicas de Análisis de Datos

47

Motivación

- Los consumidores son cada vez más complejos y exigentes.
- Tienen preferencias, hábitos y opiniones demasiado específicas y sus patrones de conducta y consumo muchas veces no son identificables mediante el uso de métodos tradicionales de análisis de información.
- La evolución y masificación de las tecnologías de información han permitido explorar nuevos mecanismos de análisis de datos de información asociada a objetivos de marketing.
- Muchas de las nuevas técnicas de análisis de datos están vinculadas al Data Mining.

48

Análisis de Datos de Marketing

- Aplicación de análisis estadístico y técnicas de modelamiento a conjuntos de datos computarizados a nivel individual para apoyar el desarrollo de programas de marketing efectivos de comunicación directa con clientes bien identificados
- Proceso que envuelve la recolección de información de clientes pasados, actuales y potenciales para la construcción de una base de datos que mejore el desempeño de los esfuerzos de marketing.
- Almacenamiento de respuesta y otra información de los clientes (estilos de vida, transacciones, etc.) en una base de datos computacional como base para programas de lealtad de largo plazo de los consumidores, para facilitar futuros contactos y para la planificación de todas las políticas de marketing.

49

Minería de Datos

- Conjunto de técnicas utilizadas para la búsqueda de información en bases de datos.
- Actividad de extracción cuyo objetivo es el de descubrir hechos contenidos en las bases de datos.
- Proceso analítico diseñado para explorar grandes cantidades de datos (generalmente datos secundarios) con el objetivo de detectar patrones de comportamiento consistentes o relaciones entre las diferentes variables para aplicarlos a nuevos conjuntos de datos.

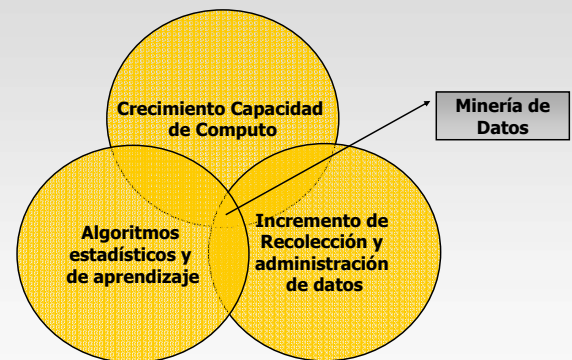
50

Razones de la Evolución de Nuevas Técnicas de Análisis de Datos

1. **Cambio del Rol del Marketing Directo.**
 - Movimiento en las relaciones de marketing.
 - Declive de efectividad de métodos tradicionales.
 - Miopía de los canales de venta.
 - Deseo de medir el impacto de los esfuerzos de marketing.
2. **Cambios en la estructura de costos.**
 - Disminución de costos de procesamiento electrónico.
 - Aumento de costos de marketing.
3. **Cambios Tecnológicos.**
 - Aparición de nuevas formas de compra y pago.
 - Desarrollo de métodos de diferenciación de las comunicaciones con el consumidor.
4. **Cambio en sociales y económicos**
 - Fraccionamiento de mercados.
 - Aumento del valor del tiempo para los usuarios.
 - Aparición de la sociedad de la información.
 - Aumento de competitividad.

51

Convergencia de 3 tecnologías claves



52

Crecimiento capacidad de computo

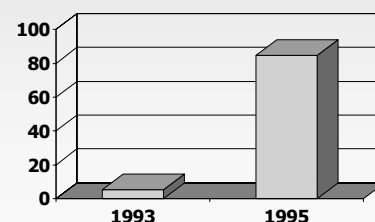
- Ley de Moore:
 - La capacidad de computo se duplica cada 18 meses.
- Hoy en día es muy fácil y barato conseguir máquinas MUY poderosas.

53

Incremento de Recolección y administración de datos

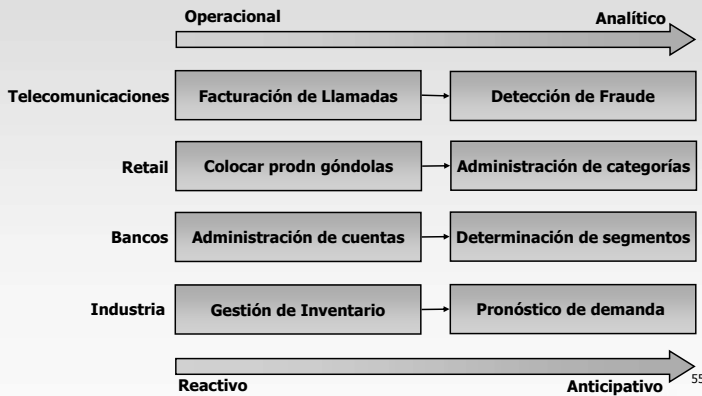
- En general, entre mas datos mejor.

% CIOs Empresas Construyendo Data Warehouse



54

Motivaciones para el almacenamiento de datos

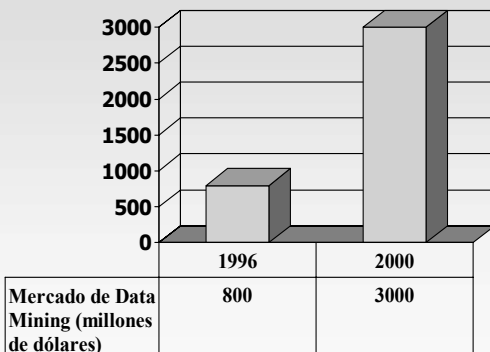


Algoritmos estadísticos y de aprendizaje

- Muchas técnicas han estado esperando por el desarrollo de la tecnología para poder ser aplicadas.
- Los estadísticos ya hacían minería de datos desde hace tiempo, pero de manera menos automatizada.
- Una buena máquina de aprendizaje es solo una aplicación inteligente de procedimientos estadísticos.
- Gran parte de la investigación en minería de datos se enfoca en el fortalecimiento de técnicas existentes para obtener pequeños porcentajes de mejora.

56

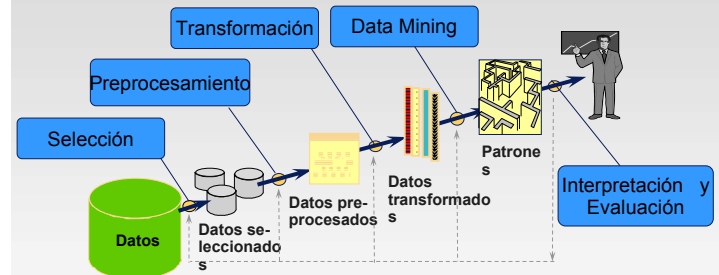
Relevancia de Data Mining



Fuente: META Group

57

KDD: Knowledge Discovery in Databases



"KDD es el proceso no-trivial de identificar patrones previamente desconocidos, válidos, nuevos, potencialmente útiles y comprensibles dentro de los datos"

58

Selección

- Se debe definir cuáles serán las fuentes de información que nutrirán los modelos de minería de datos:
 - Evaluar:
 - La relación que tiene la información contenida con los objetivos de la investigación.
 - La disponibilidad de los datos y la facilidad con que pueden actualizarse.
 - La calidad de los datos contenidos en la fuente.
 - Uso de DataWarehouse como fuente principal de información.
- El resultado del proceso de selección corresponde a una especificación de:
 - Tablas a considerar y su origen (idealmente agregar un mecanismo formal de carga).
 - Registros a considerar:
 - Filas: Definir los filtros que correspondan.
 - Columnas: Definir los campos que correspondan.

59

Preprocesamiento

- Completar valores faltantes.
 - Uso de promedios.
 - Uso de valores NULL.
- Registros duplicados.
 - Analizar si la duplicación tiene sentido o corresponde a un error.
- Datos incoherentes.
 - Ej1: Edades negativas.
 - Ej2: Una persona que mide 1.75 cm.
- Eliminación de basura
 - En los procesos de migración de las bases de datos se genera muchos datos sin sentido.
- Estandarización del contenido de los registros.
 - Fuente 1: Mujer=M, Hombre=H.
 - Fuente 2: Mujer=1, Hombre=0.

60

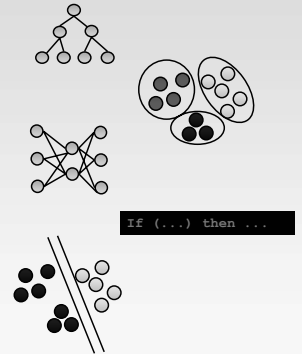
Transformación

- Dependiendo el método y el objetivo, el formato original en el que está almacenada la información no es el adecuado y deben ser transformado:
 - Transformación de datos ordinales en categóricos.
 - *Compras -> Buen cliente, Regular y malo.*
 - Introducción de variables binarias.
 - *EsLunes, EsMartes, EsMiercoles, EsfinDeSemana...*
 - Series de tiempo: A cada registro se le crean las variables $x_{t-1}, x_{t-2}, \dots, x_{t-k}$.
- Se busca dejar los datos tal como lo requieren los métodos.

61

Data Mining

- Árboles de decisión.
- Fuzzy-means.
- Redes neuronales.
- Reglas de asociación.
- Support Vector Machines.
- Etc.



62

Interpretación y evaluación

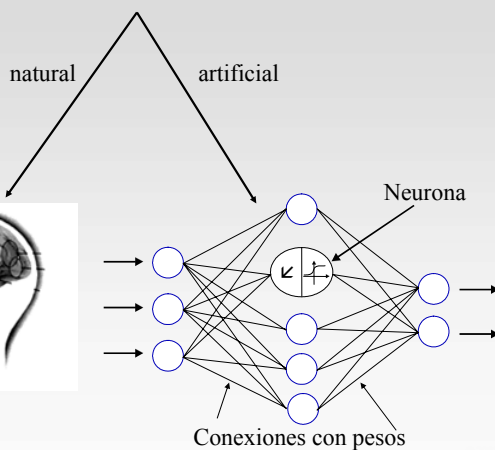
- Los resultados generados por los modelos de minería de datos deben ser interpretados:
 - Deben tener sentido para los tomadores de decisiones.
 - *Ningún tomador de decisiones le creará ciegamente a un modelo de minería de datos.*
 - La información debe ser útil para la empresa.
 - *La información generada debe ser traducible a una serie de medidas operacionales.*
 - La información debe ser nueva.
 - *Una alta proporción de la gente que compra bebidas cola, compra pisco los viernes y sábados por la noche.*
- Los modelos requieren de la incorporación de inteligencia de negocio por parte de los usuarios.

63

Redes Neuronales

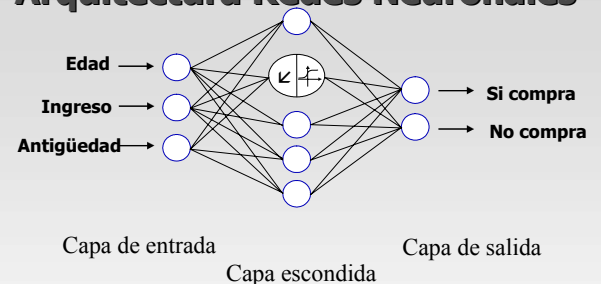
64

Redes Neuronales



65

Arquitectura Redes Neuronales



- Los inputs son transformados por una red simple de nodos de procesos.
- Los nodos se combinan (con distintos pesos) y producen una salida.
- Preguntas obvias:
 - ¿Que funciones de transformación usamos en cada nodo?
 - ¿Cómo determinamos los pesos de cada nodo?

66

Reglas de asociación

67

Reglas de Asociación: Confidence

- "Confidence":
 - La regla $X \Rightarrow Y$ tiene "confidence" c si $c\%$ de las transacciones en T con X también contienen Y .
- Ejemplo:
 - Productos : $\{1, 2, 3, 4, 5\}$
 - Transacciones : $T = \{(1, 3, 4), (2, 3, 5), (1, 2, 3, 5), (2, 5)\}$
 - $(2, 3) \Rightarrow (5)$ tiene "confidence" 100%
 - (2 de 2 transacciones que contienen (2, 3) también contienen (5))

68

Reglas de Asociación: Support

- "Support":
 - La regla $X \Rightarrow Y$ tiene "support" s en el conjunto de transacciones D si $s\%$ de las transacciones en T contienen $(X \cup Y)$.
- Ejemplo:
 - Productos : $\{1, 2, 3, 4, 5\}$
 - Transacciones : $T = \{(1, 3, 4), (2, 3, 5), (1, 2, 3, 5), (2, 5)\}$
 - $(2, 3) \Rightarrow (5)$ tiene "support" 50%
 - (2 de 4 transacciones de T contienen (2, 3, 5))

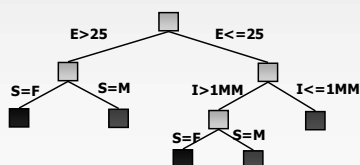
69

Árboles de decisión

70

Árboles de decisión: motivación

- **Idea:** Dada una clasificación de objetos se busca determinar una estructura jerárquica de reglas que me permitan discriminar de la mejor manera posible entre las categorías.
- **Ejemplo:** En las bases de datos de la empresa tengo una gran información de clientes y una categorización de si son Receptivos o no a la realización de una campaña de promoción.



■ Receptivo
■ No Receptivo

Los hombres mayores de 25 no serán receptivos

71

Árboles de decisión: fundamentos de los algoritmos

- Ideas básica:
 - En cada nodo del árbol elegir la regla que mejor discrimine entre las clases (maximizar la bondad de separación).
 - La discriminación es sólo sobre los objetos pertenecientes al nodo (Se parte con el conjunto completo de objetos).
 - Por simplicidad de cálculo, en cada nodo la búsqueda se hace sobre reglas univariadas.
 - El proceso se repite hasta generar nodos relativamente homogéneos.
 - Implementaciones reales suelen relajar el criterio para mejorar los tiempos de respuesta (por ejemplo un 75% de clase dominante en el nodo).

72

Aplicaciones de Data Base Marketing

73

Aplicaciones de Data Mining

- Segmentación de clientes
- Detección de Fraude
- Predicción de series de tiempo

74

Segmentación de clientes

- Optimizar esfuerzos comerciales investigando clientes con perfiles heterogéneos o con probabilidades de respuesta diferentes ante estímulos de marketing:
 - ⇒ *Promoción*
 - ⇒ *Publicidad*
 - ⇒ *Diseño de envase*
 - ⇒ *Diseño de góndola en supermercado*
- Detección de fuga: buscar perfiles con alta probabilidad de pérdida de cliente (aplicación de acciones focalizadas).
- Cross Selling: buscar perfiles con alta probabilidad que compren un nuevo producto (dado que ya consumen uno).

75

Detección de fraude

- El comportamiento de los datos en operaciones fraudulentas sigue un patrón particular.
- Dado el comportamiento de operaciones fraudulentas anteriores, se busca detectar nuevas operaciones.
- Aplicaciones:
 - Energía.
 - Telecomunicaciones.
 - Tarjetas de crédito.

⇒ **•Consumos anormales:**
•Respecto a comportamiento anterior.
•Respecto a pares similares.

76

Predicción de series de tiempo

- Serie de tiempo:
 - Serie de valores de la misma naturaleza recolectados a lo largo del tiempo:
 - ⇒ *Precios de acciones.*
 - ⇒ *Inversión publicitaria.*
 - ⇒ *Demanda.*
- Se busca estimar los próximos valores de la serie.
- Especialmente útil cuando se tiene muchos datos con una relación poco estudiada.

77

Ejemplos

78

Análisis de Respuesta con Redes Neuronales

- Promoción de nuevo servicio en un banco.
- Objetivos alternativos:
 - Maximizar la tasa de respuesta (#clientes/#mails enviados).
 - Minimizar el volumen del mailing.
 - Maximizar el número de clientes en una selección determinada.

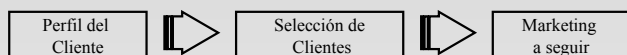
79

Análisis de Respuesta con Redes Neuronales

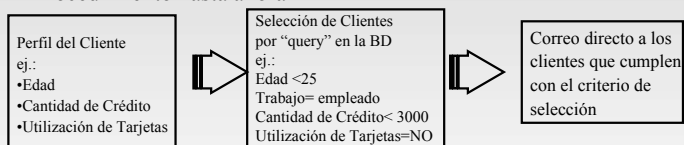
- Datos disponibles: 9000 registros de clientes de un subsegmento:
 - Uso de 18 productos.
 - Datos socio-demográficos y personales.
 - Datos de respuesta después de un mailing.
- Almacenar el resultado de esta campaña para aprender y aplicar a próximo mailing.

80

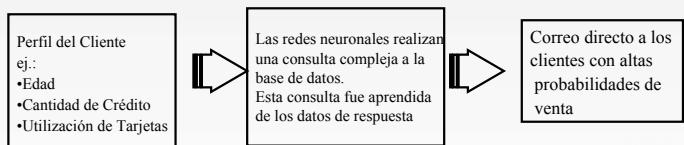
Análisis de Respuesta con Redes Neuronales



Procedimiento hasta ahora

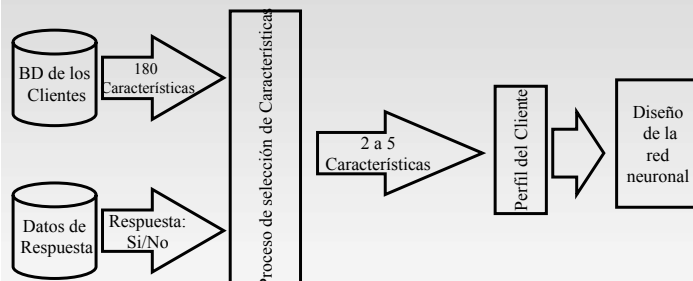


Procedimiento Futuro



81

Análisis de Respuesta con Redes Neuronales



82

Resultados

Tasa de Respuesta

- Antes de la aplicación de minería de datos.
 - 7-9%
- Después de la aplicación de minería de datos.
 - 17-20%



83

Detección de Fraude

- Caso Nexus: Sistema para detección de fraude usando árboles de decisión.
 - Memoria Rodrigo León
- Otras aplicaciones:
 - Telefónicas.
 - Isapres.
 - Pago de impuestos.

84

Relas de Asociación Caso: Supermercado

- Dick's Supermarket: cadena de supermercados en EE.UU.
 - Utilización herramientas de data mining para extraer información acerca del historial de compra de sus clientes.
 - 90% de las ventas son capturadas por tarjetas de registro
 - Data mining encontró relaciones entre productos y patrones de compras de los clientes (por ejemplo: se venden yoghurt y barras de chocolate juntos).
 - Diseñar góndolas con ambos productos aumenta ventas por 60%

85

Reglas de Asociación Caso: Wen Minning

- $\{ /compania/productos/, /compania/productos/producto1.html \} \Rightarrow \{ /compania/productos/producto2.html \}$
- $s = 0.01, c = 0.75$:
 - 75% de los clientes que accesan la sección de productos y visitan producto1.html, también visitan producto2.html, y la combinación de estos eventos ocurre en el 1% de las sesiones.

86

Nuevas Técnicas de Análisis de Datos

- Sin embargo no todas las nuevas técnicas de análisis de datos con aplicaciones de marketing provienen del mundo del Data Mining.
- Algunas provienen de la estadística (econometría):
 - Ejemplo: Clase latente, nuevo método de segmentación
- Algunas provienen de la optimización:
 - Ejemplo: DEA, nuevo método de administración de categorías en supermercados
- Tendremos dos charlas con aplicaciones de estos nuevos métodos de análisis de datos.

87